DOCUMENT RESUME

ED 458 281 TM 033 445

AUTHOR	Carr-Hill, Roy
TITLE	Monitoring the Performance of Educational Programmes in Developing Countries. Education Research, Serial No. 37.
INSTITUTION	Department for International Development, London (England).
ISBN	ISBN-1-86192-224-8
PUB DATE	1999-00-00
NOTE	195p.; Produced with Mike Hopkins, Abby Riddell, and John Lintott.
AVAILABLE FROM	Education Department, Department for International Development, 94 Victoria Street, London SW1E 5JL, England.
PUB TYPE	Reports - Evaluative (142)
EDRS PRICE	MF01/PC08 Plus Postage.
DESCRIPTORS	Case Studies; *Comparative Analysis; *Developing Nations; Elementary Secondary Education; *Evaluation Methods; Foreign Countries; *Performance Based Assessment; *Program Evaluation; Social Influences
IDENTIFIERS	*Educational Indicators; *Monitoring

ABSTRACT

The purpose of this monograph is to lay the groundwork for developing a series of indicators for education that can be used to monitory progress in education projects, in country-specific education systems, in developmental spinoffs from investment in education, and in terms of poverty reduction. The first chapter sketches a conceptual framework that analyzes the reasons for the resurgence of interest in educational performance indicators, identifies the problems of definition and development, and reviews literature about the use of performance indicators. This chapter also discusses possible frameworks for performance indicators drawn from the experience of selected countries and contexts. Chapter 2 presents case studies of experience using educational indicators in Kenya, Andhra Pradesh (India), and South Africa. Chapter 3 moves beyond the education sector to develop a framework of overall social indicators. The chapter concludes that the basic problem remains the comparability and coverage of data that are meant to be the basis for the indicators. The overall message of the report is that, while anyone can develop performance indicators, the problem is to identify the social forces that have led to the generation of data and then to take into account the misuses to which they can be put by arbitrary authority. (Contains 252 references.) (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

DFID Department for International Development

TM

ED 458 281

EDUCATION RESEARCH

MONITORING THE PERFORMANCE OF EDUCATIONAL PROGRAMMES IN DEVELOPING COUNTRIES

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

S. Arthur

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Roy Carr-Hill
with Mike Hopkins, Abby Riddell, John Lintott

Serial No. 37

TM033445

BEST COPY AVAILABLE

Department For International Development

2

EDUCATION RESEARCH

MONITORING THE PERFORMANCE OF EDUCATIONAL PROGRAMMES IN DEVELOPING COUNTRIES

Roy Carr-Hill
with Mike Hopkins, Abby Riddell, John Lintott

1999

Serial No. 37
ISBN: 1 86192 224 8

Department For International Development

DEPARTMENT FOR INTERNATIONAL DEVELOPMENT

EDUCATION PAPERS

This is one of a series of Education Papers issued from time to time by the Education Department of the Department For International Development. Each paper represents a study or piece of commissioned research on some aspect of education and training in developing countries. Most of the studies were undertaken in order to provide informed judgements from which policy decisions could be drawn, but in each case it has become apparent that the material produced would be of interest to a wider audience, particularly those whose work focuses on developing countries.

Each paper is numbered serially, and further copies can be obtained through the DFID's Education Department, 94 Victoria Street, London SW1E 5JL, subject to availability. A full list appears overleaf.

Although these papers are issued by the DFID, the views expressed in them are entirely those of the authors and do not necessarily represent the DFID's own policies or views. Any discussion of their content should therefore be addressed to the authors and not to the DFID.

No. 1 SCHOOL EFFECTIVENESS IN DEVELOPING COUNTRIES: A SUMMARY OF THE RESEARCH EVIDENCE. D Pennycuick (1993) ISBN: 0 90250 061 9

No. 2 EDUCATIONAL COST-BENEFIT ANALYSIS. J Hough (1993) ISBN: 0 90250 062 7

No. 3 REDUCING THE COST OF TECHNICAL AND VOCATIONAL EDUCATION.
L Gray, M Fletcher, P Foster, M King, A M Warrender (1993) ISBN: 0 90250 063 5

No. 4 REPORT ON READING ENGLISH IN PRIMARY SCHOOLS IN MALAWI.
E Williams (1993) Out of Print – Available on CD-Rom and DFID website

No. 5 REPORT ON READING ENGLISH IN PRIMARY SCHOOLS IN ZAMBIA.
E Williams (1993) Out of Print – Available on CD-Rom and DFID website

See also No. 24, which updates and synthesises No's 4 and 5.

No. 6 EDUCATION AND DEVELOPMENT: THE ISSUES AND THE EVIDENCE.
K Lewin (1993) ISBN: 0 90250 066 X

No. 7 PLANNING AND FINANCING SUSTAINABLE EDUCATION SYSTEMS IN SUB-SAHARAN AFRICA.
P Penrose (1993) ISBN: 0 90250 067 8

No. 8 Not allocated

No. 9 FACTORS AFFECTING FEMALE PARTICIPATION IN EDUCATION IN SEVEN DEVELOPING COUNTRIES. C Brock, N Cammish (1991) (revised 1997). ISBN: 1 86192 065 2

No. 10 USING LITERACY: A NEW APPROACH TO POST-LITERACY METHODS.
A Rogers (1994) Out of Print – Available on CD-ROM and DFID website. Updated and reissued on No 29.

No. 11 EDUCATION AND TRAINING FOR THE INFORMAL SECTOR.
K King, S McGrath, F Leach, R Carr-Hill (1994) ISBN: 1 86192 090 3

No. 12 MULTI-GRADE TEACHING: A REVIEW OF RESEARCH AND PRACTICE.
A Little (1995) ISBN: 0 90250 058 9

No. 13 DISTANCE EDUCATION IN ENGINEERING FOR DEVELOPING COUNTRIES.
T Bilham, R Gilmour (1995) Out of Print – Available on CD-ROM and DFID website.

No. 14 HEALTH & HIV/AIDS EDUCATION IN PRIMARY & SECONDARY SCHOOLS IN AFRICA & ASIA.
E Barnett, K de Koning, V Francis (1995) ISBN: 0 90250 069 4

No. 15 LABOUR MARKET SIGNALS & INDICATORS.
L Gray, AM Warrender, P Davies, G Hurley, C Manton (1995) Out of Print – Available on CD-ROM And DFID website.

No. 16 IN-SERVICE SUPPORT FOR A TECHNOLOGICAL APPROACH TO SCIENCE EDUCATION.
F Lubben, R Campbell, B Dlamini (1995) ISBN: 0 90250 071 6

No. 17 ACTION RESEARCH REPORT ON "REFLECT" METHOD OF TEACHING LITERACY.
D Archer, S Cottingham (1996) ISBN: 0 90250 072 4

No. 18 THE EDUCATION AND TRAINING OF ARTISANS FOR THE INFORMAL SECTOR IN TANZANIA.
D Kent, P Mushi (1996) ISBN: 0 90250 074 0

No. 19 GENDER, EDUCATION AND DEVELOPMENT - A PARTIALLY ANNOTATED AND SELECTIVE BIBLIOGRAPHY. C Brock, N Cammish (1997) Out of Print – Available on CD-ROM and DFID website.

No. 20 CONTEXTUALISING TEACHING AND LEARNING IN RURAL PRIMARY SCHOOLS USING AGRICULTURAL EXPERIENCE. P Taylor, A Mulhall (Vols 1 & 2) (1997) Vol 1 ISBN: 1 861920 45 8 Vol 2 ISBN: 1 86192 050 4

No. 21 GENDER AND SCHOOL ACHIEVEMENT IN THE CARIBBEAN. P Kutnick, V Jules, A Layne (1997) ISBN: 1 86192 080 6

No. 22 SCHOOL-BASED UNDERSTANDING OF HUMAN RIGHTS IN FOUR COUNTRIES: A COMMONWEALTH STUDY. R Bourne, J Gundara, A Dev, N Ratsoma, M Rukanda, A Smith, U Birthistle (1997) ISBN: 1 86192 095 4

No. 23 GIRLS AND BASIC EDUCATION: A CULTURAL ENQUIRY. D Stephens (1998) ISBN: 1 86192 036 9

No. 24 INVESTIGATING BILINGUAL LITERACY: EVIDENCE FROM MALAWI AND ZAMBIA. E Williams (1998) ISBN: 1 86192 041 5

No. 25 PROMOTING GIRLS' EDUCATION IN AFRICA. N Swainson, S Bendera, R Gordan, E Kadzamira (1998) ISBN: 1 86192 046 6

No. 26 GETTING BOOKS TO SCHOOL PUPILS IN AFRICA. D Rosenberg, W Amaral, C Odini, T Radebe, A Sidibé (1998) ISBN: 1 86192 051 2

No. 27 COST SHARING IN EDUCATION. P Penrose (1998) ISBN: 1 86192 056 3

No. 28 VOCATIONAL EDUCATION AND TRAINING IN TANZANIA AND ZIMBABWE IN THE CONTEXT OF ECONOMIC REFORM. P Bennell (1999) ISBN: 1 86192 061 X

No. 29 RE-DEFINING POST-LITERACY IN A CHANGING WORLD. A Rogers, B Maddox, J Millican, K Newell Jones, U Papen, A Robinson-Pant (1999) ISBN: 1 86192 069 5

No. 30 IN SERVICE FOR TEACHER DEVELOPMENT IN SUB-SAHARAN AFRICA. M Monk (1999) ISBN: 1 86192 074 1

No. 31 PRODUCTION OF LOCALLY GENERATED TRAINING MATERIALS. I Carter (1999) ISBN: 1 86192 079 2

No. 32 SECTOR WIDE APPROACHES TO EDUCATION. M Ratcliffe, M Macrae (1999) ISBN: 1 86192 131 4

No. 33 DISTANCE EDUCATION PRACTICE: TRAINING & REWARDING AUTHORS. H Perraton, C Creed (1999) ISBN: 1 86192 136 5

No. 34 THE EFFECTIVENESS OF TEACHER RESOURCE CENTRE STRATEGY. Ed. G Knamiller, G Fairhurst (1999) ISBN: 1 86192 141 1

No. 35 EVALUATING IMPACTS (OF EDUCATION PROJECTS & PROGRAMMES). Ed. V McKay, C Treffgarne (1999) ISBN: 1 86192 191 8

No. 36 AFRICAN JOURNALS – A SURVEY OF THEIR USAGE IN AFRICAN UNIVERSITY LIBRARIES. A Alemna, V Chifwepa, D Rosenberg (1999) ISBN: 1 86192 157 8

No. 38 TOWARDS RESPONSIVE SCHOOLS – SUPPORTING BETTER SCHOOLING FOR DISADVANTAGED CHILDREN (case studies from Save the Children). M Molteno, K Ogadhoh, E Cain, B Crumpton (2000) ISBN: **to be confirmed**

No. 39 PRELIMINARY INVESTIGATION OF THE ABUSE OF GIRLS IN ZIMBABWEAN JUNIOR SECONDARY SCHOOLS. F Leach, P Machankanja with J Mandoga (2000)
ISBN: **to be confirmed**

No. 40 THE IMPACT OF TRAINING ON WOMEN'S MICRO-ENTERPRISE DEVELOPMENT
F Leach, S Abdulla, H Appleton, J el-Bushra, N Cardenas, K Kebede, V Lewis, S Sitaram (2000)
ISBN: **to be confirmed**

NOW AVAILABLE – CD-ROM containing full texts of Papers 1-33

OTHER DFID EDUCATION STUDIES ALSO AVAILABLE

REDRESSING GENDER INEQUALITIES IN EDUCATION. N Swainson (1995)

FACTORS AFFECTING GIRLS' ACCESS TO SCHOOLING IN NIGER. S Wynd (1995)

EDUCATION FOR RECONSTRUCTION. D Phillips, N Arnhold, J Bekker, N Kersh, E McLeish (1996)

AFRICAN JOURNAL DISTRIBUTION PROGRAMME: EVALUATION OF 1994 PILOT PROJECT. D Rosenberg (1996)

A MODEL OF BEST PRACTICE AT LORETO DAY SCHOOL, SEALDAH, CALCUTTA. T Jessop (1998)

All available free of charge from DFID Education Department, 94 Victoria Street, London SW1E 5JL

CONTENTS

CONTENTS i

LIST OF TABLES AND FIGURES iv

LIST OF ACRONYMS AND ABBREVIATIONS v

EXECUTIVE SUMMARY vii

CHAPTER ONE: PERFORMANCE INDICATORS - THE CONCEPTUAL FRAMEWORK 1

1.1 Historical Interest in Performance Indicators and Lessons 2
1.1.1 The revised code in the Nineteenth Century 2
1.1.2 Earlier international attempts 4
1.1.3 The new managerialism 6
1.1.4 Resurgence of performance indicators in education systems 8

1.2 Defining and Developing Performance Indicators 11
1.2.1 Approaches to definitions 11
1.2.2 Developing useful indicators 15
1.2.3 Who should choose and design the system? 17

1.3 Uses and Abuses of Performance Measurement 18
1.3.1 Uses 18
1.3.2 Abuses or perverse uses 21
1.3.3 The (limited) value of performance measurement 23

1.4 Experiences of Other Agencies/Countries with Systems of Performance Indicators 25
1.4.1 Australia 25
1.4.2 Sweden 26
1.4.3 Commonwealth Secretariat 26
1.4.4 World Bank 27

1.5 The Specificity of Educational Systems in Developing Countries 29
1.5.1 The Jomtien Agenda 29
1.5.2 Covering the diversified field of education 30

1.6 Possible Frameworks for Performance Indicators 33
1.6.1 The DFID context 33
1.6.2 A skeleton for Frameworks 36
1.6.3 Developing indicators at different levels and for different stages 37

ANNEX 1A: Problems of Measurement at the Sectoral Level: Examples of Indicators and their Associated Problems 41

ANNEX 1B: Problems of Measurement at the Sectoral Level: Examples of Indicators and their Associated Problems 43

ANNEX 1C: Collecting data for Individual Performance Indicators 47

CHAPTER TWO: CASE STUDIES 49

2.1 Introduction 50
2.1.1 The choice of case studies 50
2.1.2 The approaches to evaluation 50

2.2 Measuring the Impact of Education in Kenya 51
2.2.1 Introduction 51
2.2.2 Education and training in Kenya 51
2.2.3 VET in Kenya 53
2.2.4 Kenyan data 54
2.2.5 DFID in Kenya 58
2.2.6 Concluding remarks 61

2.3 Performance Indicators in the Andhra Pradesh District Primary Education Project 63
2.3.1 The objectives of the project 63
2.3.2 Quantitative and qualitative studies: project planning process 64
2.3.3 Qualitative case studies 65
2.3.4 Indicators of implementation 67
2.3.5 The overall effectiveness of the project 71
2.3.6 Conclusion 77

2.4 Education Performance Indicators in a New South Africa: A Progress Report 77
2.4.1 Framework: Levels and purposes of use of performance indicators 78
2.4.2 Current educational initiatives 82
2.4.3 Indicator development and use 84
2.4.4 Conclusion 92

ANNEX 2A: Further Proposals for the Development of Educational Performance Indicators in South Africa 94

CHAPTER THREE: BEYOND EDUCATIONAL PERFORMANCE INDICATORS: MOVING TOWARDS OVERALL SOCIAL INDICATORS 96

3.1 Introduction 97

3.2 Background 97
3.2.1 Origins and purposes of systems of socio-economic data 97
3.2.2 The social indicators movement of the 1960s 98
3.2.3 Approaches: a thousand flowers bloom 100
3.2.4 Factors contributing to the decline of social indicators 101

3.3 Social Indicators Systems 101
3.3.1 Early examples in the seventies 101
3.3.2. Basic needs approach 103
3.3.3 Moving towards the individual 105
3.3.4 Conclusion 106

3.4 Composite Quality of Life Indices 108
3.4.1 The human development index (HDI) 108
3.4.2 Other innovatory composites 110
3.4.3 Conclusion 111

3.5 Developing a Modern Framework: Monitoring Social Conditions Top-down? 112
3.5.1 A framework? 112
3.5.2 Designing more relevant statistics? 113
3.5.3 An excess of estimates, data drought 119

3.6 The Impact of Education upon Health 120

3.6.1 Introduction 120
3.6.2 Brief review of evidence of possible impacts 120
3.6.3 The measurement problem in evaluating projects 122
3.6.4 Conclusion 123

3.7 The Impact of Education on Employment in Developing Countries 124
3.7.1 Current practice 124
3.7.2 What improvements could be made in indicators? 127

3.8 Monitoring Basic Needs At the Local Level 129
3.8.1 Some basic principles of indicator construction for democratic monitoring 129
3.8.2 An appropriate local information monitoring system 131

ANNEX 3A: Opportunity Cost And Valuations 134

ANNEX 3B: Technical Problems In Developing Indicators 141

ANNEX 3C: Valuing Health And Measuring (Un)Employment 147

ANNEX 3D: Exercises In Local Planning 152

CONSOLIDATED REFERENCE LIST 157

LIST OF TABLES

Table 1:	Different Purposes of Performance Indicators	20
Table 2:	Collecting Data about Non-Formal Education	33
Table 3:	The Planning Process and the Need for Performance Indicators at Different Levels	38
Table 4:	Performance Indicators at Different Stages of the Planning Process and Project Cycle	40
Table 5:	Annual Educational Census Data	55
Table 6:	Education and Employment Indicator Availability	57
Table 7:	Generic Indicator Framework Applied to Kenya	62
Table 8:	Project Specific Indicators for the Participation Objective in the APDPEP	74
Table 9:	Matrix of Actors and Purposes of Performance Indicators in Education in South Africa	80
Table 10:	A Possible Framework of Social Concerns	107

LIST OF FIGURES

Figure 1: Extract from Logical Framework used for SPRED II
Figure 2: Types of Performance Indicators Used by Different Agencies
Figure 3: Performance Indicators Needed By Different Levels of Management
Figure 4: Quantitative and Qualitative Studies: Project Planning Process

ACRONYMS AND ABBREVIATIONS

APDPEP	Andhra Pradesh District Primary Education Programme
APPEP	Andhra Pradesh Primary Education Programme
AwHF	Adjustment with a Human Face
BLS	Bureau of Labour Statistics (USA)
BRC	Block Resources Committee
BSSRS	British Society for Social Responsibility in Science
CIDA	Canadian International Development Agency
DALY	Disability Adjusted Life Years
DEOs	District Education Officers
DFID	Department for International Development
DPEP	District Primary Education Programme
DPO	District Planning Office
ECCE	Early Childhood Care and Education
ECSIP	Eastern Cape School Improvement Project
ELET	English Language Educational Trust
ELT	English Language Training
EMIS	Education Management Information System
EEC	European Economic Community
EUROSTAT	European Union Statistics Office
GDP	Gross Development Product
GER	Gross Enrolment Ratio
GIS	Geographical Information System
GNP	Gross National Product
HDI	Human Development Index
HMI	Her Majesty's Inspectors
HEDCOM	Heads of Education Departments Committee
ICLS	International Conference of Labour Statisticians
IEA	International Association for the Evaluation of Educational Achievement
IEQ	Improving Educational Quality
IIEP	International Institute for Educational Planning
ILO	International Labour Office
INES	International Network of Education Systems
ISEW	Index of Sustainable Economic Welfare
IT	Information Technology
KAB(P)	Knowledge, Attitudes, Behaviours (Practice)
KCPE	Kenyan Certificate of Primary Education
KCSE	Kenyan Certificate of Secondary Education
KNEC	Kenyan National Education College
LDC	Less Developed Countries
LMIS	Labour Market Information Systems
LSMS	Living Standards Measurement Study
MELS	Meeting of Experts in Labour Statistics
MoE	Ministry of Education
MIS	Management Information Systems
MRA	Manpower Requirements Approach
MRP	Mediterranean Regional Project
NASA	National Aeronautic Space Agency
NFRHRD	National Foundation for Research into Human Resources Development
NGO	Non Governmental Organisation
NNW	Net National Welfare
NQF	National Qualifications Framework (South Africa)
ODA	Overseas Development Administration
OECD	Organisation for Economic Co-operation and Development
OFSTED	Office for Standards in Education
ORPIs	Outcome Related Performance Indicators
PCA	Participatory Community Appraisal
PERSAL	Personnel Information System (South Africa)
PETRA	Primary English Teaching for Rural Areas (South Africa)
PIs	Performance Indicators
PMIS	Project Management Information System

PPS	Probability Proportional to Size
PQLI	Physical Quality of Life
PRISM	Primary Schools Management Project (Kenya)
PSPIs	Programme Specific Performance Indicators
PTA	Parent-Teacher Association
PTR	Pupil-Teacher Ratio
QALY	Quality Adjusted Life Year
RDI	Relative Deprivation Indicator
ROR	Rate of Return Analysis
RRI	Rate of Reduction of Insufficiencies
SAQA	South African Qualifications Authority
SCERT	State Centre for Educational Research and Training (India)
SEEA	Satellite Systems for Integrated Environmental and Economic Accounting
SIDA	Swedish International Development Cooperation Agency
SPRED	Strengthening of Primary Education
SSDS	System of Social and Demographic Statistics
TACs	Teacher Advisory Centres
TCs	Teacher Centres
UK	United Kingdom
UNDP	United National Development Programme
UNESCO	United National Educational, Scientific and Cultural Organisation
UNICEF	United Nations International Children's Emergency Fund
UNRISD	United Nations Research Institute for Social Development
UNFPA	United Nations Fund for Population Activities
USAID	United States Agency for International Development
VECs	Village Education Committees
VET	Vocational Education and Training
WMS	Welfare Monitoring and Evaluation Survey (South Africa)
WTP	Willingness to Pay

EXECUTIVE SUMMARY

The overall purpose of this monograph is to lay the groundwork for developing a series of indicators for education that can be used to monitor progress in education projects, in country specific education systems, in developmental spin-offs from investment in education and in terms of poverty reduction. In the current policy climate, the focus is on basic education.

In **Chapter One**, a 'conceptual framework' is sketched out. This includes, in the first three sections, an analysis of the reasons for the resurgence of interest in educational performance indicators, identifying the problems of definition and development and reviewing the literature about the use and abuse of performance indicators.

We show that performance indicators and their critics are not at all new. The issue is not whether performance indicators are good or bad, but what questions are being asked to which performance indicators might provide an answer. The most generic definition is preferred: information that can be used for understanding, and eventually for decision-making. The potentially distorting effects of too rigid a system of performance indicators are identified in terms of seven characteristics observed in the management literature: tunnel vision, sub-optimisation, myopia, convergence, ossification, gaming and misrepresentation.

The second half of Chapter One discusses possible frameworks for performance indicators drawn from the experience of a selection of countries and contexts. Key questions are identified as:

- Is the performance indicator about a significant aspect of the education system or the impact of education?
- Can it be readily understood by everyone involved both in-country as well as by external parties?
- Will the data be reliable and not subject to significant modification as a result of response error, or changes in the personnel generating it?
- To what extent is the data reported under the control of operational managers?

The apparent similarity of the problems in different DFID programme countries and the similarity of the solutions proposed by 'international experts' would suggest that there could be agreement on a set of indicators. Indeed, it is not technically difficult; but, insofar as partnership and collaboration with developing countries themselves are valued, then the appropriate indicators should be defined through a process of negotiation, not *a priori*.

While concrete sets of indicators are not developed, a framework is proposed based on distinctions between:

- context, aims, inputs, processes, outputs and outcomes;
- the range of possible stakeholders; and
- types and levels of decision-making.

This framework can be used at the sectoral level based on DFID's overall aims; at the planning and pre-planning stage; and at the project implementation and monitoring stage. Detailed specification of the indicators within this framework should be seen as a collaborative effort.

In **Chapter Two**, case studies of the experience in Kenya (with a long-standing DFID involvement in various projects), Andhra Pradesh (where there has been a large scale unified programme for over six years), and South Africa (where appropriate structures are being developed) are examined.

The Kenyan example shows that, despite several decades of project involvement in Kenya, there is little understanding as to exactly what the Government gets out of spending nearly 30 per cent of its recurrent budget on education; and little movement towards some basic monitoring.

The Andhra Pradesh case study was based on the experience of designing a 'participatory' monitoring and evaluation scheme. Experience with the previous project had shown that a top-down scheme of monitoring and evaluation was of only limited utility. The problem was, therefore, to design mechanisms for collecting data at several different levels that would allow the construction of 'appropriate' performance indicators immediately useful for local project management. We demonstrate that, in this situation, the usual distinction between monitoring and evaluation breaks down; and the possibility of indicators is strictly limited by the constraint of identifying simple yet robust data collection techniques.

The South African example shows the difficulty of developing sets of performance indicators at the same time as appropriate structures for the education system.

In **Chapter Three**, we move beyond the education sector to develop a framework of overall social indicators. The rise of what has been called the 'social indicator movement' in the 1960s is discussed, drawing attention to the major split between those focusing on a uniform method of valuation (usually money) across the social sectors and those concerned to reflect the diversity of living patterns. The concerns that led to the development of social indicators in the 1960s continue to be relevant today. Examples of different approaches to developing social indicators systems are reviewed.

We conclude that the basic problem remains the comparability and coverage of data that are meant to be the basis for the indicators. The experience both in the

1960s and now is that composite indices based on combining different data hide more than they reveal.

While recognising that it is time consuming, we recommend celebrating diversity in the approach to indicator development. The final two sections consider, therefore, the different kinds of problems that arise when attempting to develop a modern framework for monitoring social conditions top down and for monitoring the satisfaction of basic needs at the local level.

The overall message of the report is that whilst anyone can develop performance indicators, the problem is to identify the social forces which have led to the generation of data, and therefore to take into account the misuses to which they can be put by arbitrary authority.

CHAPTER ONE

PERFORMANCE INDICATORS: CONCEPTUAL FRAMEWORK

CHAPTER ONE

PERFORMANCE INDICATORS: THE CONCEPTUAL FRAMEWORK

This chapter is organised into six sections. Section one presents a schematic history of performance indicators in developed countries. The definition and development of performance indicators are discussed in section two. Section three examines the uses and abuses of performance indicators. Lessons gained from the experiences of development agencies with educational performance indicators are reviewed in section four. Section five looks at the current educational context in developing countries in relation to DFID's position. Section six proposes an appropriate framework for developing performance indicators.

1.1 LESSONS FROM PREVIOUS ATTEMPTS TO USE PERFORMANCE INDICATORS

This section selects examples of educational monitoring from the past 150 years to illustrate crucial points for later development of the theme.

1.1.1 The Revised Code in the Nineteenth Century

In response to Benthamite/utilitarian ideas of efficiency, early Victorian England discovered examinations. From 1846, pupil teachers had been paid following the results of an annual examination. The first Code for elementary schools was introduced in 1860: to obtain a government grant, the school needed to abide by the Code. In 1861 the Newcastle Commission reported on the state of elementary education in England and Wales saying "They leave school, they go to work, and in the course of a year they know nothing at all". From 1862, under the Revised Code, more popularly known as 'payment by results', the bulk of the nation's elementary school pupils were subjected to an annual examination. The schools were paid four shillings a year for the satisfactory attendance of each pupil between the ages of six and 12, and eight shillings for each pupil, dependent upon the results of an annual examination, not restricted by age but by standard, carried out by Her Majesty's Inspectors (HMI).

The Code had some positive achievements: a reduction in public expenditure on education; attendance at elementary schools certainly improved and basic literacy was extended. Elementary teachers, however, became increasingly dependent on market forces and their success rate in the annual examination (Hogg, 1990:9). One HMI reported:

> The tendency of the new Code is to cause managers and teachers to regard simply the pecuniary grants, and all that does not tend to produce an increased result as to these is hardly taken into account ... The expression on a child's failure to pass any subject is not to regret at his ignorance so much as indignation at his stupidity and the consequent loss. (Committee of Privy Council on Education Report, 1864: 114)

Reviewing this history, Hogg suggests that the issue is not whether Performance Indicators are good or bad, but what questions are being asked to which performance indicators might provide answers. When using Performance Indicators to assess schools, for example, the questions underlying the exercise might be: What is education for? What difficulties in the education system is it thought performance monitoring might improve? How far will monitoring and assessment be successful in meeting present and new needs? (Hogg, 1990:12).

Teachers' reactions to the Code were virulent:

> A young teacher was dying of consumption, but on hearing of the death of the inspector he started up, a wild light struck out of his eyes, like fire from steel and he said with a hideous broken scream: 'By God, I hope he's in hell!" (Runciman, 1887:29)
>
> ...sooner than teach in an elementary school, under any one of a score of inspectors I could name, I would go before the mast in a collier, or break stones on a casual ward - or better die! (Runciman, 1887: 29)

These quotes highlight how an inappropriately designed set of performance indicators can have profound effects on those directly affected by them. This suggests that it would be wise to check the likely impact of a set of Performance Indicators with those most directly affected by them before full implementation.

During the late nineteenth century and early twentieth century, more scientific methods of testing school children began to be developed in the United States, France and the United Kingdom; and were introduced in the 1920s. While there were critics, particularly among teachers, testing was widespread between the two World Wars, although confidence in assessment began to wane from the mid 1930s.

Looking at more recent history and the growth of performance indicators in the educational field, Hogg (1990:7) suggests that it was the launching of the Soviet Sputnik in 1958 and the American concern to 'catch-up', as demonstrated by the American National Defense Education Act, that led to the revival of interest in the use of performance indicators in the 1960s. This was then absorbed into the more general 'social indicator movement' attempting to monitor the impact of technological change (à la Bauer, 1966). With the oil crisis and the retrenchment of the public sector (including central statistical agencies) in Western capitalist countries, there was a hiatus until the revival in the United States with the publication of *A Nation at Risk* (National Commission 1983). This coincided with other tendencies (see section 1.4) to generate the current level of world-wide interest. Is it perhaps appropriate to talk of long waves of performance indicators?

1.1.2 Earlier International Attempts

International Association for the Evaluation of Educational Achievement

The International Association for the Evaluation of Educational Achievement (IEA) has been in existence for over thirty years; and has been carrying through a testing programme since the beginning of the 1970s. Their first efforts were limited to testing in mathematics, because it was presumed at the time that the test material in that subject area was the least likely to suffer from cultural variation.[1] Since then there has been testing of language, mathematics and science in over 25 countries, and in several countries more than once; although the issue of cultural and systemic variation between countries (Cummings and Riddell 1992) is still unanswered.

The school focus of the IEA, however, is insufficient in countries where not everyone in an age group goes to school. This is acknowledged by IEA: "Surveys of formal schooling alone are not sufficient to assess education in third world countries, if one wants to judge and compare an entire age or grade population cohort." (Plomp and Loxley 1992) The particularities of developing performance indicators in the area(s) of non-formal education are considered further in section 1.5.

Organisation for Economic Co-operation and Development (OECD): A Quarter Century of Educational Performance Indicators

A very early publication was *Indicators of Performance of Education Systems* (Carr-Hill and Magnussen, 1973) which suggested the following framework for thinking about the goals of education systems for which indicators needed to be developed.

- transmission of knowledge and skills;
- education and economy;
- equality and educational opportunity;
- provision of educational services for individual requirements; and
- education and the quality of life.

Carr-Hill and Magnussen argued that as no clear set of educational goals is available, nor can be imposed (from which indicators can be derived), differing goals should be treated separately. For example, the different outlooks of the pedagogue (the learning process), of the economist (education as an investment

[1] Although that was already being debated - see Pateman, (1968). Subsequently the whole sub-discipline of ethno-mathematics has developed (see, for example, Bishop 1997).

and supplying qualified manpower), of the sociologist (concerned with access to education and the effect on social class stratification), of the 'consumer' (estimates of private demand for education) and of the philosopher (questions of the impact of education on the quality of life) should all be considered separately. Although their recommendations did not lead to a regular publication of Performance Indicators in education, a set of 45 indicators covering some aspects of this framework were adopted in principle. (see *A Framework for Educational Indicators to Guide Government Decisions*, OECD, 1974).

Most subsequent authors have implicitly agreed with the liberal eclecticism of this approach; but literally dozens of different frameworks have been proposed (van Herpen 1992). The type of system actually proposed by the OECD in its second incarnation is of greater interest to developing countries - if only because of its relative simplicity. Following on from two international conferences - Washington (1988) and Poitiers (1989) - the system was initially organised around three areas:

- the economic, social and demographic context of education;
- cost, resources and processes; and
- the results of education.

In later editions of *Education at a Glance* (various years since 1993), four subdivisions are used:

- the demographic, social and economic context of education;
- the costs of education and human and financial resources;
- participation in the educational process; and
- the outcomes of education.

The section on participation in the educational process is further subdivided into:

- access to education, participation and progression; and
- school environment and school/classroom processes.

The section on the outcomes of education is further subdivided into:

- graduate output of educational institutions;
- student achievement and adult literacy; and

- labour market outcomes of education.

A number of working groups were formed to develop appropriate indicators for each of these sub-sections: educational outcomes; student destinations; school features and processes; and expectations and attitudes to education. They continue to meet under the aegis of OECD.

1.1.3 The New Managerialism

While we have demonstrated that management by objectives in the form of payment by results is not new, undoubtedly there has been a greater emphasis in recent years on outcomes. It is important to recognise that this is part of an overall trend, especially in the public sector, associated with the move away from direct state provision, and probably with the globalisation of competition consequent upon the alleged triumph of Western capitalism. While it is not suggested that this cryptic analysis should be explored in depth, the examination of the reasons for this greater emphasis might give us some clues as to what works and what does not work.

The Recent Growth of Performance Indicators as a Management Tool[2]

There have been performance indicators in the UK public sector for some time, but they have tended to focus on the managerial process, the principal objective (Carter, 1989) being to enable closer control of devolved management by central government. The introduction of outcome-related performance indicators (ORPIs) was intended to enhance accountability to external interested parties - service users, taxpayers, or auditors acting on their behalf. Anthony and Young (1984: 649) argue that, in non-profit organisations, ORPIs are addressing the most important stimulus to improved management control, i.e. 'more active interest in the effective and efficient functioning of the organisation by its governing board'. Increased sensitivity of representatives to popular preference will then permeate the organisation and - as in the private corporate sector - will have a profound impact on internal control mechanisms.

Thus, throughout the private sector, control has effectively been exercised through the financial accounts - what we could call 'input accountability'. The presumption in the private sector has been that, in competitive product markets, consumers can observe directly the merits of competing goods and services - so the private sector tends[3] to pay little attention in accounting terms to the quality of the output, although of course they may advertise the quality of their products to attract the consumer. But most public services - even after decentralisation and/or quasi-privatisation - are effective local monopolies, so the citizen cannot directly

[2] Some of the arguments in this sub-section rely heavily on Smith (1993).

[3] The increased interest in environmental and social responsibility accounting may signal a change in the private sector; see Hopkins 1999.

experience the services provided (or value-for-money) in different localities; and many citizens may not directly experience, for example, the police and fire services even though they value them. One role of ORPIs, therefore, is to act as a proxy for the direct experience of services provided by alternative jurisdictions - i.e. to address a concern for geographical and social equity.

Assuming also that taxpayers not only wish to see tax revenues being used well but also that they believe they can exert some influence, performance indicator systems are one means of communicating relevant information. Here, the analogy with the shareholder in the private sector is closer as the concern is with efficiency and effectiveness. The use of performance indicators for input accountability is-based on the familiar principal -agent model of management (Balman, 1990). Public sector managers, however, are responsible in varying degrees to a much wider variety of constituencies than the single 'principal' envisioned in this model.

Essentially, therefore, the growth of performance indicators is linked to a demand for accountability and equity, and demands to demonstrate value for money in respect of activities, in the context of evident differences between providers as a consequence of decentralisation. Moreover, "on the assumption that well-informed electors will not tolerate manifestly inefficient management teams, it can be argued that performance systems will also encourage managerial efficiency in the use of resources" (Smith, 1993:137). Equally, from the cybernetic point of view (Hofstede, 1981), an ORPI system *should* enhance the political control of the public sector by offering individual citizens timely and meaningful feedback on the effect of public sector activity, and *should* influence the design of the organisation's internal control system.

Communicating Performance Indicators

Performance indicator systems are not easy to scrutinise as the reader/user may have to disentangle at least four causes of variability in terms of:

- the objectives being pursued by different organisations;
- the social and economic environments;
- the accounting methods used; and
- the levels of (managerial) efficiency.

This has led to the argument for an expert intermediary; hence the role in the UK of the National Audit Office (with responsibility for auditing government departments including DFID) and the Audit Commission (with reference to Health and Local Authorities). They both have the remit to publish relatively easy-to-read reports on the performance/value-for-money of activities of the corresponding organisations. In the education system, the Inspectorate fulfils the 'traditional' role of monitoring educational quality from the centre and highlighting problem areas, while the Office

for Standards in Education (OFSTED) has a more public face. The latter represents a form of 'recentralisation' at least in terms of publishing league tables of school performance along a number of dimensions (not only in terms of examination results).

In developing countries, the local inspectorate is usually inadequately resourced and often dysfunctional; indeed, the improvement of the system of school supervision is returning to donor programmes (see Khaniya, 1997). Donors, of course, audit their own programmes but that usually tends to be internal. With the increase in joint funding and sector programmes, the mechanisms have become slightly more public: for example, the several donors involved with the District Primary Education Programme (DPEP) in India combine in a Joint Supervision Mission to assess progress. Such missions fulfil a role halfway between the inspectorates and public accountability.

The Importance of Evidence - and of Judgement

Excessive reliance on performance measures in a complex society where the public sector is under pressure can have unintended and often dysfunctional consequences (Merchant, 1990) mirroring those of the mid-nineteenth century commented on by Hogg (1990). To a greater or lesser extent, most data reported within a performance indicator system can be controlled by operational managers. Unless the system is a perfect reflection of all the intended inputs, processes and outputs of an organisation *from the point of view of those operational managers* (which is very unlikely), there is potential for distortion (see section 1.3.2). The problem - a major theme throughout this report - is that we need to be able to assess the impact of performance indicator systems upon the performance of the individuals, institutions or organisations (or units within organisations) whose performance is being monitored. While there are many anecdotal accounts, systematic evidence of that kind is sorely lacking.

No one denies the importance of evidence of performance: it is a *sine qua non* of evaluating any professional practice. Part of the professional role, however, is to make judgements in complicated situations. While those who promote performance indicator systems emphasise the importance of very careful interpretation, much less attention is paid to *institutionalising* the role of such interpretation and judgements.[4]

1.1.4 The Resurgence of Performance Indicators in Education Systems

The previous section focused on the growth in the use of performance indicators in

[4] Perhaps the clearest example of the attempt to recognise the interplay between evidence and judgement is in the criminal law. Both prosecution and defence try to build up a convincing picture - based on the presentation of evidence - to place before the jury. But this does not mean that lawyers or juries ignore the evidence: indeed it would be seen as rather silly to crusade for the use of evidence in criminal trials: the issue is *how* it is used.

the public sector generally: what about the particular case of education?

Ruby (1989), reviewing the Australian experience, suggests a number of reasons for renewed policy interest in indicators:

(a) a concern to improve the country's international economic competitiveness by a variety of means but particularly by increasing the general level of education of the workforce;

(b) demands by decision-makers for better information about outcomes and performance to improve policy-making about education - the 'what works' syndrome;

(c) demands for information to guide and monitor the implementation of reforms, particularly structural reforms involving the devolution of authority, and to evaluate the outcome of those reforms;

(d) political commitment to equity such as equality of outcomes for minority groups;

(e) a belief that better information about effective strategies and performance will bring about qualitative improvements in teaching and learning;

(f) enhancing accountability measures in the public sector by gathering data on performance and outcomes; and

(g) a commitment to improving the information available to the public about the performance of public authorities.

Educational Indicators and Accountability

Wyatt (1994) agrees that the concept of educational indicators as summary statistics on the status of education systems is not new. Whenever there are perceptions of falling levels of achievements, the traditional response has been a call for greater accountability and the imposition of higher 'standards'.

He cites Carley's (1981) explanation for the decline in the social indicators of the 1960s: expectations were too high regarding the time needed for their elaboration and the (in-)adequacy of social theory; eagerness to supply social indicator information often led to the provision of poor quality data thereby undermining confidence; insufficient attempts to relate indicators explicitly to policy objectives. While his commentary is rather naïve (see chapter three), these constitute important warnings about the appropriate expectations, quality of data, and clear specification of relationship between indicators and policy objectives.

Wyatt (1994) suggests that the recent pressure towards educational indicators is due to calls for accountability; and the requirement of central government for a

means of monitoring the process of devolution of responsibility to the school. The latter led to a call for school accountability in respect of centrally determined criteria; how schools might evaluate themselves emphasising the use of locally determined indicators in the school management process; and the use of indicators to monitor specific policy objectives in schools. Note that both the first and third are forms of counterbalancing 'recentralisation', following decentralisation to schools.

Commenting upon the potential use of performance indicators, Wilcox says:

> First, performance indicators are seen as an essential element in the greater accountability which will be demanded of schools as a consequence of financial delegation...[second] there is a concerted attempt...to develop appropriate [performance indicators] but also to model and interpret them (Wilcox, 1990:32).

Recent Trends

Scheerens (1992) identifies three recent trends:

- a transition from descriptive statistics (largely input and resource measures) to measurement of performance outcomes;
- a movement towards more comprehensive systems and a growing interest in manipulable characteristics;
- a concern to measure data at more than one aggregation level.

He also shows how different indicators are appropriate according to the type, level and mode of decision making:

Types of Decision Making – whether we are interested in: the organisation of instruction; the planning of education and establishing the structures within which it is delivered; personnel management; or resource allocation and use.

Levels of Decision Making - whether at the level of the school; lower intermediate authority (e.g. districts); upper intermediate authority (e.g. provinces or regions); central authority.

Modes of Decision Making - varying from full autonomy to collaborative; to independent but within a framework (although the extent to which the latter is different depends on how tight the frame is).

The moves toward Sector Wide Approaches have raised the profile of performance inddicators: thus "the shift towards strenghening primary education is a notable and welcome development but there are significant difficulties attached to SWAPs.

In paticular evaluation is not significantly integrated into design of projects". (DFID Evaluation of Primary Schooling, Synthesis Study). One could add that another significant difficulty is that the context of a sector programme includes not only the educational systems but also other social sectors.

Most evaluations at least pay lip-service to some logical framework – or equivalent management tool – constructed at the beginning of the project or programme, with specific indicators being taken as good proxy indicators for the attainment of one objective rather than another. An example, given in Figure 1, taken from the log frame proposed for Kenya's SPRED programme shows clearly how different indicators are seen as relevant to the attainment of different specific objectives.

The agencies most concerned with 'performance indicators' are unsurprisingly The World Bank and USAID with the EC/EU not far behind (see Figure 2): the World Bank rates projects on five areas USAID proposes very broad-brush indicators and the EC/EU also retains some straightforward indicators. There is little awareness of the difficulty of collecting reliable data for these indicators."

Figure 2: Types of Performance Indicators Used by Different Agencies

(A) World Bank
Evaluation carried out by Operations Evaluation Department. All projects rated According to three results oriented criteria – outcomes, sustainability, institutional Development; and two process oriented – Bank performance.

(B) USAID

- Education's share of national budget
- Primary education's share of education budget (for recurrent and capital Expenditure); and
- Share of primary recurrent, non-salary expenditure of primary budget.

As an indication of effective schools, the use of a fundamental quality and equity Level (FQEL) index, which measures the number of schools meeting minimum Criteria in services and coverage ... a means of capturing the united elements that go into making an effective school and the idea of "access-with quality (USAID 1998: 41).

(C) EC/EU
Education indicators are well known. Some of them, like those now being chosen to support structural adjustment in Burkina Faso, may be used in the context of the new conditionally approach: school-attendance rates (boys/girls); first-year primary Attendance rates (boys/girls), success rates in end-of-primary exams (boys/girls), number of books per pupil; level of satisfaction among users; cost and rate of use of informal education by adults. A gender breakdown of indicators is essential here.

Narrative Summary	Measurable Indicators	Means of Verification	Important Assumptions
GOAL: 1 Increased demand for and utilisation of high quality primary education	1.1 Reduced wastage rates, especially for girls, from current 56% to less than 30% by 2005. 1.2 Improved student performance, raising KCPE average by 20 points and stabilising repetition rate at 15% through to 2005. 1.3 Increase current GER of 79% to 85% by 2005. 1.4 Increased pupil and parental satisfaction	1.1 MoE Planning Unit statistics 1.2 KCPE examination results 1.3 MoE Planning Unit statistics 1.4 MoE/Project surveys	(Goal to Supergoal (ODA Aim 2 Statement): 1 GoK and parents continue to maintain present level of commitment to education
PURPOSE: 1 To improve the quality and cost-effectiveness of teaching and learning in primary schools on an equitable basis	1.1 Teaching and learning environment improved in all districts by 1999, though all teachers using new skills that promote active learning, and through use of textbooks provided under project. 1.2 Improved professional support and inspection service to schools nationwide through upgraded and diversified Teacher Advisory Centre (TAC) system and upgraded inspectorate, by 2000. 1.3 Strengthened community participation in school management and delivery: i) through effective school committees established in all schools by 1997 ii) community education programmes operational and sustained in at least 5 disadvantaged districts by 2001. iii) information, education and communication programmes operational nationwide (with special emphasis in all areas of high wastage and low enrolment in first 3 years of project) by 2001. 1.4 MoE use operational research to improve resource allocation and planning: i) ensuring al severly disadvantaged schools receive free textbooks by 2001. ii) teacher recruitment and deployment systems improved based on reduced student staff ratios, and piloting of multi-grade teaching by 2001. iii) rationalised cost-sharing system in place in all districts by 2000. Reduction by at least 10% in parental expenditure per child in schools in receipt of project textbooks.	1.1 Impact assessments. Inspectorate and TAC Tutor reports. 1.2 District Education Board reports, Inspectorate Reports, project reports. 1.3 Impact assessments, consultants reports, District Education Board reports. District Education Officer reports. 1.4 MoE statistics, planning and budgeting reports, District Education Officer reports, consultants reports.	(Purpose to Goal): 1 Effective co-ordination with PRISM project 2 Non-ODA funded components achieve targets 3 Effective targeting of project resources through operational research to districts where wastage and non-enrolment are greatest. 4 Consistency of KCPE examinations 5 No increase in poverty in targeted districts.
OUTPUTS: Component 1: 1 Improved teacher training through School-based in-service Teacher Development Programme (STD)	1.1 STD operational in all target districts by 1999 and minimum of 90% (14,000 schools) coverage nationwide by 2001. 1.2 Teachers guides prepared for circulation to 175,000 teachers by mid 1998. 1.3 School advisers guides prepared by mid 1997. 1300 TAC tutor trained in management of STD and 16,000 mentor teachers trained by 2000. 1.4 Testing and Measurement booklets prepared for circulation to 175,000 teachers by KCPE examiners by end of 1997.	1.1 Project impact study, Inspectorate reports, consultants reports. 1.2 Project reports, consultants reports 1.3 Project reports, consultants reports, TAC tutor reports, Inspectorate surveys. 1.4 Project reports, consultants reports	(Output to Purpose): 1 Orientation of headteachers to STD through PRISM project 2 Politicians and unions accept the desirability of improved student:teacher ratios and the concept of multi-grade teaching in remote rural areas. 3 Institutional strengthening components received broad-based support 4 Policy commitment to increasing non-salary components of primary budget is retained 5 Teachers able to apply new teaching methodologies.

Why Have Performance Indicators Now Become Acceptable?

The devolution of governance to districts and to schools has meant that they have become interested in comparing their performance with that of others. In turn this has stimulated discussion over what should count as performance. This was not the case in the 1960s and 1970s when the first systematic systems were proposed (e.g. OECD 1973). Unless a central authority has the power (as in nineteenth century England), it is very unlikely that a performance indicator would be acceptable, let alone implemented, without convergence on what is meant by good and bad performance.

Bottani and Delfau (1992), introducing the OECD project, argue that the influences on the development of an indicator system are: policy considerations; research knowledge; technical considerations; practical considerations; and the (usually political) 'choosers'. These principles interact and sometimes conflict, although policy relevance is likely to be the major driver. Fewer indicators rather than more are likely to be preferred.

For DFID, however, the arguments of Scheerens about the kinds of indicators required for different levels, modes and types are important: what types of decision (choosing new projects, monitoring on-going projects)?; at what level (head office, country, project)?; in what mode (in-house, with the recipient country, for the general public)? To this one might add distinguishing between the kinds of data required for different educational outcomes. This is developed in the final section where we discuss a possible framework.

Finally, we must not forget that the basis for any indicator system is the quality of the basic data (whether qualitative or quantitative) that are collected. This depends on the extent to which the field officers (in this case teachers) accept the system and can collect the appropriate data with the willing participation of the community.

1.2 DEFINING AND DEVELOPING PERFORMANCE INDICATORS

1.2.1 Approaches to Definitions

There are as many different 'concepts' of educational performance indicators cited in the literature as there are different dissections that can be made of an educational system. The lists complied by Ashenden (1987a) and Sing (1990), for example, cite a range of indicators of effectiveness, equity, productivity, process, quality, among many others.

The sources of data for performance indicators are eclectic: the data provided by the institutions which are part of administrative information systems; the data based on client and provider perceptions collected by questionnaire; or information collected through direct observation of the workings of the institutions (Wilcox, 1990:37).

The most frequently cited definition of performance indicators in the educational context is that of Oakes (1986:1-2). He argues that these indicators must provide at least one of the following kinds of information:

- a description of performance in achieving desired educational conditions and outcomes;
- features known through research to be linked with desired outcomes;
- a description of central features of the system in order to understand its functioning;
- information which is problem oriented; and
- policy-relevant information.

Of course, Oakes is writing mainly about school or college-based education in developed countries, contexts to which his approach can be applied relatively easily. It is much more difficult, however, to describe context and performance for out-of-school provision, or for the recruitment into school of the disadvantaged, in developing countries (see section 1.5).

He goes on to argue that indicators should be:

- ubiquitous features of schooling found in some form throughout the systems/settings being compared;
- enduring features of the systems so that trends over time can be analysed;
- readily understandable by a broad audience;
- feasible in terms of time, cost and expertise;
- generally accepted as valid and reliable statistics.

Again, in the context of developing countries, that is not always easy.

For the objectives of this report, we propose the most generic definition of performance indicators possible: **information that is useful for understanding levels and variations in performance, in order to assess the impact of interventions and ultimately inform decision-making.**

As Wyatt says, there is a certain consensus that performance indicators are "statistics which reveal something about the health or performance of education, describe its core features and are useful for decision-making" (Wyatt, 1992:106). The problem at that level – performance indicators for an education system - is the necessity to build a coherent set that provides a valid representation of the

condition of the education system. The problem here - for the assessment of DFID education projects and sector performance in those countries where DFID is involved - is the necessity to build consensus with project partners before defining the indicators (see DFID 1997, *White Paper*). Hence our reluctance to define those indicators which should not be defined *a priori*, but through a process of agreement.

1.2.2 Developing Useful Indicators

The shared basis of the approaches is an input-process-output model (van Herpen, 1992). The apparent reduction of the complexity of educational systems and their interaction with other societal systems is, however, illusory. Not only can an output of one activity easily be an input to another, an output from the perspective of one person or group can be seen as an input by another. Equally, the environment is sometimes seen as explanatory (having influence on the outcomes of the educational process), while some focus on the impact education has on the economic and social environment (the outcomes). Even if these problems are ignored, there are still problems of moving from a set of agreed (verbal) values to agreed indicators, identifying the target audience for a particular set of indicators, and then choosing the most appropriate data.

Problems of Moving from Agreed Values to Agreed Indicators

There are fundamental differences in the cultural significance and the social interpretation placed on agreed values, and in the willingness to identify and quantify a policy problem. Ruby (1992) argues that the choice of indicators should reflect either - or preferably both - common policy problems or the enduring values that underpin education systems (Nuttall, 1992).

Socio-cultural issues, however, not only mediate the demand for the collection of specific data series, they may also affect the way in which indicator systems are developed overall. For example, the nature of institutional linkages, the absence of political will, shortcomings in technical capacity, and different time frames will mediate between the basic values and interests shaping national or system policies and the priorities that are reflected in the data systems that are developed.[5]

These socio-cultural differences between institutions or societies can be very important. The obvious example is the disjunction between the time frames of data collection and policy. Data systems (as distinct from data *collection)* can show change over time; but whilst policy concerns emerge quickly and demand

[5] Alach and Carr-Hill (1989: 29) provide a concrete example of how countries vary in relating data to policy in respect of the debates over inequalities in health. They point to: the extent to which the political regime in power is prepared to recognise the problem; the extent to which the problem can be documented; prevailing views about causation; the particular form of system via which services are delivered; the economic and historical context; and the relative position and power of disadvantaged groups.

immediate answers, data systems are costly to change. Pragmatic agreement over a (sub) set of indicators is a possible way of relieving this tension. Account needs to be taken, however, of the different philosophies about the hierarchies of policy and action (see chapter two for a discussion about this in context of South Africa).

Identifying the Target Audience

Equally important, different professions and disciplines have different perspectives that come to bear on the agreement process. For example, in caricature, researchers belonging to the effective school movement focus on process indicators; economists on the relationship between input and output; political scientists on the possibilities of steering the system; sociologists focus on the environment; and teachers on survival.

For policymakers, good information is *simple, comparable and timely*. To technicians, however, this often means something different. For example, technical comparability might emphasise common definitions and collection times; whereas, for the policymaker, the rationale for comparability is in order to trade off or exclude some options by reference to experience elsewhere or previously (without their necessarily being full comparability in a technical sense).

Along similar lines, politicians want information that is *accessible, direct and public*; (Riley and Nutall, 1994). For example: the choice of schools has been a stimulus to demands for information; but, previously, this may not have been an issue because, from the point of view of the parents, there was no easy way to assess the difference between schools and hence no reason to choose.

One response to multiple audiences and ambiguity of purpose is to collect more information (Stern and Hall, 1987); but this is not necessarily a solution. On the one hand there is the danger of 'information overload' (see the case study of Andhra Pradesh in chapter two); and, on the other, of the demand for pseudo-scientific indices to reduce the confusion (see chapter three, section three).

What Data to Collect?

In technical terms, a good indicator is relevant, reliable, understandable, and can be updated. These requirements are not always easy to fulfil. Moreover, the different uses to which performance indicators are put reflect different data gathering requirements:

- Some provide a benchmark for measuring change over time; others are focused on differences across geographic areas or institutions at a point in time.

- Some reflect a policy issue, or an aspect of education that might be altered by a policy decision; others information relevant to managerial processes.

- Some are macro and quantitative, reflecting broad-brush decisions and others are micro and qualitative as part of a change process.

1.2.3 Who Should Choose and Design the System?

The issues of how indicators are chosen and by whom has itself generated a large literature focusing, in particular, on policy, technical and practical considerations. Broad agreement exists on the need for valid, reliable, timely, comparable, feasible, reasonably costed, policy relevant and comprehensible indicators (Nuttall 1992:93). There remain differences, however, about "the number, the need for redundancy, and the extent to which the indicators should be comprehensive and organised by and into a framework that reflects the functioning of the education system with...known causal links." (Nuttall 1992:93)

The Policy Agenda

Choices are inevitably influenced by value systems of those making the choice. McDonnell suggests that:

> The policy context then plays two distinct roles in the design of a system of indicators. First, it provides the major rationale for developing and operating such a system. Second, the policy context constitutes a key component of any educational indicator system, because specific policies can change the major domains of schooling in ways which affect educational outcomes (1989:241-242).

Nuttall (1994), however, emphasises the importance of creating indicators that are independent of the current policy agenda otherwise it will be difficult to maintain a stable statistical system.'

A Model of the System

Several authors argue that a set of Performance Indicators must reflect scientific understanding of how the education system functions, but it also needs to reflect the interests of the policy making community, including consumers and data producers. Again Nuttall cautions:

> ...the present understanding of the educational process is insufficient for the postulation of a [precise] model, but that it is possible to create a framework that embodies the available limited knowledge of empirical relationships and that begins to relate malleable variables to desirable outcomes without promising too much. (Nuttall, 1994:85)

Criteria for choosing and developing and evaluating indicators

Nuttall suggests a series of important practical lessons that have been learnt from the various attempts to develop performance indicators in developed countries:

- indicators are diagnostic and suggestive of alternatives rather than judgements;
- any implicit model [however partial] must be made explicit and acknowledged;
- criteria for selection must be clear and related to an underlying model;
- individual indicators should be valid, reliable and useful, etc.;
- comparisons must be done fairly and in a variety of different ways; and
- the various consumers of information have to be educated about what the indicators mean, how they are to be interpreted and what consequences they might have.

For any proposed system of monitoring education programmes or projects in developing countries, the objective should not be to provide a comprehensive, causally specified framework. Instead, it would be useful to have small sets of indicators organised roughly into the categories of inputs, context, processes and outputs (perhaps distinguishing between levels and modes of decision-making and kinds and types of data). At the same time, the danger of keeping the indicator set too small, and of corrupting the behaviour of those whose performance is being monitored, needs to be recognised (Darling-Hammond, 1992).

1.3. USES AND ABUSES OF PERFORMANCE MEASUREMENT

Despite a reasonably thorough search of the literature, and although much has been written on both the advantages and disadvantages of performance indicators, it has been difficult to find anything relevant to the public sector[6] other than theoretical accounts.

1.3.1 Uses

Sizer, Bormans and Spee (1992) identify five core uses of performance indicators in government institutional relationships: monitoring, evaluation, dialogue, rationalisation, and resource allocation. Where performance indicators become more controversial is where the emphasis shifts from their use as one of the many inputs into effective decision-making, to using them as a ranking device to allocate

[6] As Smith (1993) demonstrates, within the private sector the task is much easier because of the focus on financial inputs and outputs.

esteem and funding differentially (Meek, 1995).

For Oakes (1986), indicators should be used to:

- report the status of schooling;
- monitor changes over time;
- explain the causes of various conditions and changes;
- predict likely changes in the future;
- profile the strengths and weaknesses of the system;
- inform policy makers of the most effective way to improve the system;
- inform decision making and management; and
- define educational objectives.

In the context of development, the World Bank (1996) suggests the following range of uses:

- clarification of the objectives and logic underlying the strategic plan;
- promotion of efficient use of resources via performance accounting;
- forecasting and early warning during program implementation;
- measuring programme results for accountability, programme marketing and public relations;
- benchmarking in order to learn from success; and
- measuring customer (beneficiary) satisfaction for quality management.

In the DFID context, the most important uses are probably for accountability, marketing and public relations.

Audience and purpose

The problem is: do policymakers, politicians and the public want the same information for the same reasons? Table 1 below shows the different uses of performance indicators from the perspective of different audiences. Unless, there is agreement over the objectives and uses of performance indicators, then the PI

reports will never be accepted.

Table 1: Different Purposes of Performance Indicators

Audience	Purpose
Institutional	Internal Comparisons with others Marketing Evaluation of teachers
Government	Accountability Policy and planning Allocation Overall level of funding Value of investment Manpower planning
Public	Accountability
Student	Choice
Teachers	Self-assessment
Industry	Research funding Graduate employment

Source: taken from Davis 1996 Report for Commonwealth Secretariat

Concerns about the Development and Use of Performance Indicators

Davis (1996) summarises the concerns about the use of performance indicators as follows:

- costs of additional data;
- emphasis on one aspect;
- inappropriate when institutions have different objectives;
- their use in isolation;
- loss of diversity;
- imposition of central control;
- limited value for quality; and
- effectiveness and efficiency emphasised more than quality.

Compare this with the list of problems enumerated previously in an internal document by ODA (ODA 1995):

- indicators in isolation may mislead;
- simple models mask reality;
- collection and analysis of indicator data can be difficult; and
- indicators may be massaged for other purposes.

In the latter list, the issues about costs and diversity are not mentioned at all; and the limitations in terms of monitoring quality are not registered. Without this kind of understanding, any indicator system that is developed is likely to have over-ambitious aims.

Moreover, there is a salutary warning from the management science literature reflecting on the way most of the performance indicator systems used in the UK have been imposed by central government with little consultation. As Jones (1986) notes, the predisposition amongst operational managers to indulge in dysfunctional behaviour is likely to be heightened if they perceive that a control mechanism is being imposed on them against their will. This increases the possibility that any private gains to be made from distorting behaviour will be exploited by at least some managers. Evidence from the private sector indicates that the style with which control schemes are implemented may have a profound impact on their effectiveness (Hopwood, 1972; Otley, 1978; Kenis, 1979).

DFID should clearly take great care in the way in which the systems are developed. Whilst the argument for bottom-up, appropriate participatory approaches is a separate and, of course, very important issue in development, the crucial importance of involving country partners is central to the above discussion (see DFID 1997, *White Paper*).

1.3.2 Abuses or perverse uses

> For every performance indicator, questions must be asked about the implied message, the behavioural implications. In other words, knowing that certain indicators are being collected and monitored, what implications do people draw? (Fitz-Gibbon, 1990:2)

Smith (1993) enumerates seven ways in which excessive use of outcome-related performance indicators might influence public sector managerial behaviour:

Tunnel vision: Concentration on areas included in the outcome-related performance indicator scheme to the exclusion of other important areas.

Suboptimization: The pursuit by managers of their own narrow objectives, at the expense of strategic co-ordination.

Myopia: Concentration on short term issues to the exclusion of long-term criteria, which may only show up in outcome –related performance indicators in many years' time.

Convergence: An emphasis on not being exposed as an outlier on any outcome-related performance indicator, rather than a desire to be outstanding.

Ossification: A disinclination to experiment with new and innovative methods.

Gaming: Altering behaviour so as to obtain strategic advantage.

Misrepresentation: Including 'creative' accounting and fraud.

On the basis of a small number of interviews with managers in the health sector; Smith (1993) suggests that all of these are possible effects, but the most likely problems were with 'gaming' and 'misrepresentation'.

The opportunity for 'gaming' in the public sector is exaggerated by the poor understanding of most production functions, illustrating the importance of choosing outcome-related performance indicators (ORPIs) with great caution, and of ensuring that incentives are compatible with organisational objectives.

Moreover, because outcome-related performance indicators are thought to require 'expert' interpretation, which tends to be provided by the manager responsible, there is considerable scope for interpretative 'misrepresentation'. Whether such misrepresentation occurs to any great extent is of course a matter of conjecture. As Smith's interviews suggested an acute awareness that other units might indulge in creative interpretation, one must conclude that there is a strong possibility that it exists. This is compounded when, as in the public sector, there are very few outside 'experts' able to give reasonably dispassionate commentary on performance measures.

Abuse of Educational Performance Indicators

In the education sector, one can see the opportunity for each and every one of the seven processes described above.

Tunnel vision: If head-teachers are only rewarded for those examined then there will be a tendency to focus teaching only on those children to be examined to the neglect of others.

Suboptimization: If head-teachers are rewarded according to the number of children in the school, there will be a tendency to devote energies to 'capturing' new pupils even though this does not improve attainment.

Myopia: An example of short-termism is where in-service training of teachers currently in post is neglected for cost reasons.

Convergence: the fear of being an 'outlier' might assume particular significance in a corporatist environment where either the state or the unions are powerful.

Ossification: New curricula and innovative teaching methods are likely to absorb staff time and be disliked because of that.

Gaming: Although unlikely to be a problem in developing countries, the reaction of head-teachers to league tables is a good example.

Misrepresentations: Where head-teachers are rewarded by numbers of children on the school register, there will be a strong temptation to falsify registers.

Smith (1993) concludes that most public sector performance indicator schemes have been developed on the assumption that they are neutral reporting devices, and too little attention has been given to the organisational context in which they will be used. The management control literature suggests that such a cavalier attitude to context threatens the objectives of the scheme.

1.3.3 The (Limited) Value of Performance Measurement

Whilst such abuses or perverse uses may be rare, the limitations have to be understood

Limitations of Performance Measurement

There are constraints in transforming theoretical concepts of outcomes into practicable measurement procedures, thus:

> "The gap between our academic aims and available measures is important because, to the extent that educational indicators have direct consequences attached to them, as in the case of performance indicators, these limited measures begin to reform classroom practice in their image. There is an assumption that policy action based on indicators will produce a desired result. Indicators are intended to advance constructive action, but such action is contextually embedded. Variations in culture and basic understanding about the inter-relationships of individuals, family, school and society are features of educational systems. The imminent danger is that the indicator model will frame the subsequent discussion in essence becoming the implicit model for schooling everywhere." (Bryk and Hermanson, 1993)

These arguments are well understood in terms of 'teaching to the test'. Performance indicators provide a useful focus on achievement, but top-down

approaches aimed at using testing to bring about change are limited unless linked to support for school improvement" (Selden, 45:1994). The argument, however, is more general. Flamholtz (157:1983) notes that 'an accounting system cannot be viewed as a control system *per se*; rather [it] must be seen as a part of a carefully designed total system of organisational control'. As Hofstede (1981:200) notes: 'the more formalised a control system, the greater the risk of obtaining pseudo-control rather than control'. Any measurement of performance needs to be introduced for a system and not just for individuals, or individual organisations within that system (Walsh, 1994); otherwise, inevitably, there are distortions.

Performance Indicators are, at most, Useful at Different Levels

The distinctions made by Scheerens (1993) between levels, modes and types of decision-making have already been discussed: and this and other distinctions made above have implications for the type of indicators that will be useful at the different levels. Both the case study of developing a monitoring and evaluation system in Andhra Pradesh and the relevance of different indicators at different levels of the new South African System (see chapter two) illustrate this well.[7]

Moreover, it needs to be emphasised that most systems restrict themselves to the 'school system' and there is relatively little attention given to adult and out-of-school education (see Carr-Hill, 1989). Van Herpen (1992) suggests that these should be built into a comprehensive system. But the information requirements of out-of-school education are very different from the formal schooling system (Carron and Carr-Hill, 1991, see section 1.5).

Care in Elaborating Performance Indicators

The lessons for DFID from these observations are that there are a number of crucial questions to ask of any proposed performance indicators:

- Is the performance indicator about a significant aspect of the education system or of the impact of education?
- Can it be readily understood by everyone involved both in-country as well as external parties?
- Will the data be reliable and not subject to significant modification as a result of response error, or changes in the personnel generating it?
- To what extent are the data being reported under the control of operational managers and therefore subject to potential distortion?

[7] See, also, the last chapter of Carron and Ngoc Chau (1996).

1.0. EXPERIENCES OF OTHER AGENCIES/COUNTRIES WITH SYSTEMS OF PERFORMANCE INDICATORS

This section reviews some of the lessons learnt by countries attempting to institutionalise school-based performance indicator systems, and relevant innovations. Hence the apparently curious choice of agencies/countries.

1.4.1 Australia

Ruby (1994) highlights six lessons that emerged from only the first twelve months of institutionalising a performance indicator system:

1. It is difficult to communicate accurately and economically about ways of assessing performance given:

   - the technically and theoretically complex nature of indicators; and,
   - any external assessment potentially challenges concepts of professionalism, traditional notions of autonomy and raises questions about the nature of accountability.

2. Technical problems of outcome measurement do not predominate; instead the focus is on problems of interpretation and the influence of contextual and process variables.

3. The importance of stressing fundamental questions of why there is a demand for performance indicators, and in what context and for what purposes they are useful, rather than on the technical and practical questions of constructing indicators.

4. Indicators are essentially normative and goal oriented, directly linked to policymaking and the political process, and only useful when linked to a model of the education system.

5. The importance of drawing on as many paradigms and perspectives as possible, to involve people working in science policy and public policy as well as education.

6. The benefits of exploring new ideas using a co-operative and relatively open process. Keeping the process of analysis transparent to those affected by the outcome establishes credibility for the outcome of the process and of those involved.

Commentary for DFID

While it is not feasible to incorporate all these lessons (e.g. to have an agreed model of every education system as a basis for a framework), issues of communication, interpretation, pluralism and transparency are important to bear in

mind when designing sets of performance indicators which will be acceptable within different country contexts.

1.4.2 Sweden

The Swedish National Agency on Education set up an evaluative project in 1992 to examine the non-cognitive development of pupils in Swedish schools. The project took account of pupils' own views as 'connoisseurs of their own schools'. It examined pupils' development on four core variables which reflected strong national purposes: independence, self-confidence, participation in decision-making and solidarity with others. This was based on the view that:

> Individuals that hold a critical mind and are used to act in independent ways are seen as important parts of the assurances that the Swedish society have taken towards fascism. (Ekholm and Karang, 1993:13)

In other words, it was important to measure self-confidence as this was seen as a prerequisite for successful learning, and involvement in decision making as essential to sustaining democracy. Tolerance and understanding of others were also seen as essential to democracy (Ekholm and Karang, 1993:14).

Commentary for DFID

The purpose of supporting primary education in developing countries is, in part, support of democratisation (Western style) - see ODA (1993) and DFID (1997). Thus, while simple counts of participation or registration in primary school - and, eventually, attendance at primary education - may have been sufficient at one stage, discussions about measures of outcome or success are usually in terms of achievements in literacy, numeracy and science. But, where democratisation is a central tenet of a donor's strategy (as with the DFID), appropriate indicators of aid effectiveness are required; at a minimum, measures of school effectiveness *have to include* non-cognitive achievements.

1.4.3 Commonwealth Secretariat

Davis (1996) set out to compare the progress made in developing performance indicators at the higher education level in Commonwealth countries. In the preface to her study, Fielden writes:

> "Surely, we thought, at the very least, we can obtain staff-student ratios from various different jurisdictions" (Fielden, preface Commonwealth Higher Education Management Service, 1996.

Despite the rhetoric about performance indicators, however, little was actually obtained, and so:

> "Our aim is to show how very little has been achieved and how, despite the massive industry of researchers working on performance indicators, comparatively few are in use nationally." (Fielden, *op cit*)

Commentary for DFID

Part of the problem was that performance indicators are used for different purposes in different contexts (see section 1.3 above). This study graphically demonstrates the gulf between the identification of a possible indicator - and even ways of collecting, relatively cheaply, the corresponding data - and the use of an indicator to inform policy.

1.4.4 World Bank

Prior to 1987, a basic data sheet giving country economic and social and sectoral data was required for all Bank-financed education projects. This was abandoned subsequently, as the data collected were often of poor quality and not seen as relevant to individual projects.

General experience of indicator use

The World Bank Operations Department report that there are:

- disparate educational information systems between and within countries;
- differences in educational systems;
- differences in classification and terminology; and
- imbalances in the collection of data.

They identify possible reasons as: complexity of education, lack of resources, lack of capacity to carry out educational research, the political nature of educational data and lack of standardisation of education system components (McRae, 1990).

The World Bank's recent approaches (or at least the rhetoric about their approaches) are increasingly characterised by a focus on developing local capacity for project benefit monitoring and evaluation. This shift of focus has occurred in response to the Wapenhans Report (1990) which points to:

- too much emphasis on the mechanics of project implementation;
- poor identification of risks and factors influencing project outcomes;

- lack of objective criteria, transparency and consistency across units; and

- ratings which tend to be overly optimistic.

For the purposes of the Bank and its clients, the most significant benefits of performance indicators accrue in project design, project supervision and monitoring and project evaluation (World Bank, 1996).

The approach is spelt out in a paper by Sigurdsson and Schweitzer (1994). This paper discusses three types of data: basic data (providing socio-economic background and context), education sector data (useful in project identification and evaluation), and project performance data (to mark the progress of project components towards specific targets).

The appropriate types of indicators related to the project cycle are: input indicators; process indicators (to monitor stages of project implementation); output indicators are the immediate project targets identified as project components to be completed; and impact indicators are derived from sectoral data (See Annex 1A for further details).

The World Bank suggests that: "Policy related indicators can be used to identify risk and enabling factors during preparation and appraisal for projects and systems." (Sigurdsson and Schweitzer, Executive Summary, 1994). It is recognised, however, that to be meaningful, education indicators must be analysed in the context of system needs and available financing. The danger that funding decisions based on indicator performance may encourage skewed or falsified data recording is also acknowledged (1994: 3).

Sigurdsson and Schweitzer conclude with the following recommendations:

- that project performance indicators be project specific;

- that a uniform approach to economic justification be applied;

- that consistent attention be paid to process indicators; and

- that the World Bank should assist UNESCO in their ongoing work to define desirable definitions for MIS systems in Bank-financed projects.

Commentary for DFID

Many of the lessons from the World Bank's experience and some of its recommendations should be adopted. Disregarding the Bank's obsession with a uniform economic justification of projects, the emphasis on project-specific performance indicators, the importance of process as well as outcome indicators and the need to pay careful attention to the source data, are all sensible points.

1.5. SPECIFICITY OF EDUCATIONAL SYSTEMS IN DEVELOPING COUNTRIES

1.5.1 The Jomtien Agenda

In some ways, the Jomtien Conference can be read as yet another attempt to implement universal primary education. While the Conference acknowledged the potential role of alternatives (at least rhetorically), the predominant view was that appropriate indicators should be based on enrolments. From the point of view of developing any indicators that go beyond simple numbers of children in schools (the simplicity is in the statement not in the counting!), the emphasis on quality is the most important.

> The focus of basic education must...be on actual learning, acquisition. It is therefore necessary to define acceptable levels of learning, acquisition for educational programmes to improve and apply systems of assessing learning achievement. (Education for All Conference, March 1990)

The Jomtien agenda focused on:

- strategies for improved training of teachers and education managers;
- alternative methods of improving access;
- increasing production and dissemination of teaching learning materials; and
- efforts at strengthening education administration, planning and management.

Potentially, of course, such a list provides the opportunity for considerable divergence over what should be done to improve quality, and therefore generates difficulties when different stakeholders try to agree on the appropriate performance indicators. As suggested in the previous section, in order for performance indicators to be developed, there has to be clarity over objectives and therefore convergence (if not consistency) between the different interested parties. According to Hoppers (1994), however, the context in which the Jomtien Agenda is being implemented has provided a fertile ground for the development of performance indicators. He points to the almost universal phenomena of:

- stagnating enrolments and an apparent deteriorating quality;
- the need for greater economies in the development aid budget;
- the internationalisation in technologies of curriculum development; and
- intensive interaction amongst policy makers promoting the same views.

The similarity of the problems and proposed solutions to which Hoppers points would suggest that there could be agreement. Ruby (1994) argues, however, that apparently similar values and policy concerns do not necessarily mean that the same indicators are relevant (see also Blunt, 1995). Given that DFID prefers developing sets of indicators in conjunction with the developing countries themselves (*White Paper*, 1997), then it will be important to allow for potentially divergent frameworks; and indeed to have a mechanism for identifying - at least internally - disagreement and non-consensus.

1.5.0 Covering the Diversified Field of Education

The analyses discussed above - because they are based on the experience in (post) industrialised countries - have all been focused only on the schools as the vehicle for education. In many poorer countries, this is not applicable: either the quality of schooling is so variable, or the main vehicle for education is outside school. The former highlights the importance of developing sensitive performance indicators, the latter is the subject of this section.

Non formal programmes have been distinguished from formal programmes along a number of dimensions. While the original distinction made by Coombs *et al* (1973) into formal/non-formal/informal based on the degree of hierarchy etc. have been shown to be inadequate, no other single set of dimensions is any more successful. As Carron and Carr-Hill (1991) show, this is because it is important to understand:

- their aims and objectives;
- the kinds of clientele they serve;
- the organising agency; and
- the relationship with the formal educational system.

They go on to distinguish four types of programme:

- para-formal or parallel educational programmes;
- professional and vocational education;
- personal development with no specific professional intent; and
- popular education.

These types are described briefly below together with the kinds of indicators that might be appropriate.

Para-formal education: The set of programmes designed for educational equivalencies to officially recognised primary, secondary or higher educational diplomas. Case studies (e.g. Bibeau 1989; Gallart 1989) have demonstrated that there has been a progressive tendency for the formal educational system to absorb 'innovations' from the non-formal education sector as part of the standard curriculum.

In addition to these second chance para-formal education programmes, there has been a rapid expansion of the private tutoring of regular formal school students. It has grown with the massification of formal education as elite middle class parents who perceive their previous privileged position to be disappearing, have sought ways of retaining the competitive edge for their children. At the same time, for formal school teachers in many developing countries, who have seen their salary eroded over the last decades, the private tutoring system has been a welcome opportunity to increase their income. Obviously, from the point of view of the parents and the tutors there are clear criteria of success; but one might also want to assess the impact of this private tutoring system upon the formal schooling system and this is more difficult.

Professional and Vocational Education: Although this is an obvious grouping, there are problems in defining the outcomes both for the individual and society. This is not the place to rehearse the well-known arguments about the difficulty of interpreting rate of return analyses (see Hough 1991), but the point is that some vocational qualifications clearly are used for screening, and some vocational education is intended to socialise people for the general 'world of work'. Moreover it seems sensible to make a distinction in vocational education and training between *general vocational education* referring to the transmission of skills, knowledge and behavioural traits which are broadly relevant to performance in all or a considerable number of occupational roles (learning to learn); and *specific instruction* which is concerned with the performance of a single task (or set of tasks) within a single job or occupation within a single institutional locale (learning to do) which is limited in scope and non-portable in application.

While these distinctions are sensible, only the latter (specific instruction) provides a basis for elaborating appropriate performance indicators from the point of view of the company in terms of improved productivity on the specific job. In principle, there will also be enhanced income and/or security for the individual so long as they remain in that job; however, the medium or long-term outcomes for the individuals may not be so successful. The arguments about the rate of return in general mean that it is not possible to define indicators for general vocational education that would be generally accepted.

Personal Development: The rapid expansion of personal development activities is one of the most significant common trends in the diversification of the educational field. This is mostly a phenomenon in the North and so may not be very relevant

here. Appropriate indicators should be reasonably clear, however, given that the purpose of this type of non-formal education is to fulfil individual wants. Data collection on client satisfaction would, therefore, be appropriate.

Popular Education: Finally, another separately identifiable example is the type of education used as a means of consciousness-raising, practised for example by the Catholic communities in Latin America during the 1980s. This model of collective promotion appears to have weakened in favour of the spectacular emergence of personal development activities but there are still situations where a liberating form of education is seen as an essential vehicle for political and social movement.

In terms of the focus of this report, however, the point is that, in this mode, education is being used as a vehicle for a totally different perspective on society. Performance here is measured in terms of revolutionary not evolutionary outcomes.

In principle, therefore, we have been able, in most cases, to identify performance indicators that would be appropriate for the different types of non-formal education. At the same time, we have to be realistic about the possibilities of collecting data. A major stumbling block is simply the 'countability' of each of the different kinds of educational activities as some non-formal education programmes do not always record enrolment. From this perspective one could separate the various strands of non-formal education into two groups:

( ) those where the providing institution would hold enrolment and/or registration data and which could therefore be captured in principle through a census survey of institutions

( ) those where the most practicable way of obtaining estimates would be through a sample population survey

These various possibilities are summarised in Table 2 below.

Table 2: Collecting Data about Non Formal Education

Level	Name	Standard	Non-standard		Suggested Classification
		State	Collectable from Institutions	Collectable via Household Surveys	
1	Pre-primary	Nursery	Playgroups	Child Care	Para-formal
2	Primary	Primary	Evening classes	Street children	Para-formal
3	Secondary	Secondary	Evening classes	Youth groups	Para-formal
4		Further Education Colleges	Industry	Youth groups, back-street colleges	TVET General/ specific
5	Tertiary	University	'Open' Universities	Auditing	Personal Development
6	Post-Doctoral	University	'Open' Universities	Auditing	Personal Development

1.6. POSSIBLE FRAMEWORKS FOR PERFORMANCE INDICATORS

The breadth and range of the above definitions and approaches means that it is probably not possible to develop a comprehensive set of indicators reflecting a definitive theoretical framework. It is possible, however, to lay out the skeleton of overlapping frameworks, which will need to be completed as appropriate.

1.6.1 The DFID Context

1. Aims of ODA Education Aid

The details of the current DFID educational strategy are being drafted. The aims set out in "Learning Opportunities for All" (DfID, 1999).

Priority is given to meeting the International Development Goals of:

- Universal Primary Education (UPE) in all countries by 2015.
- Demonstrated progress towards gender equality and the empowerment of women by eliminating gender disparity in primary and secondary education by 2005.

And, in a minor key, DfID will also help to promote adult literacy, lifelong learning and the acquisition of practical skills for development, for women and for men.

Obviously these could be the basis for a generalised indicator framework, given that, at least within DfID, there can be consensus over what is meant by UPE, and what would count as gender equality (although measuring levels of adult literacy and skill acquisition are more difficult).

The more detailed 'Framework for Action' is less clear cut. DfID will support the efforts of people and governments committed to:

Effective and Equitable UPE

- overcoming barriers to access and retention
- supporting children to complete a basic cycle of education
- improving the quality of schooling
- equity for all children
- placing UPE within the under education sector

Gender equality in school education

Literacy and Skills Development

Knowledge and Skills for Development in a Global World

Sustainable, Well Managed Education Institutions Systems and Partnerships

It is however less clear how to measure these components and sub-components. For example, considering *Effective and Equitable UPE:*

- whether access should be measured in terms of attendance or completion and therefore the identification of disadvantaged areas;
- what is meant by improvement and quality
- the most appropriate way of placing UPE within the wider education sector

The other components also pose a number of difficulties for measurement. There have been many debates over what gender equality means, there are disputes over different academic literacies, the relationship between local and global economies is contentious, and qualifying educational institutions as 'well managed' begs the question. On the other hand, it might well be possible to generate indicators if these rather general statements were better specified.

2. Planning Education Projects

The ODA guide for planning education projects (ODA 1991a, cited in Hirani Ratcliffe 1994:4) highlights several key questions:

- What is the evidence of demand for the education service proposed?
- What will be the benefits to the country and individual?
- Can these improvements be measured?
- Will there be cost economies resulting?
- Is the proposed strategy seen as the most cost-effective?
- Are the recurrent cost implications manageable?

Again, in principle, each of these questions could be used as the basis for developing a (small) set of indicators; but even then there will probably be too many. It is not clear that they could be combined into a comprehensive performance indicator framework.

3. What makes a project successful?

The conclusions of an ODA review (ODA 1993) as to what were the essential pre-requisites for a successful project included:

- a conducive policy environment;
- joint commitment to project goals and outcomes by donors and government;
- effective project design;
- local ownership including participatory appraisal;
- local financial and institutional capacity for implementation and sustainability;
- effective management and administration of donor inputs; and
- effective monitoring and reporting.

These are much vaguer: it is difficult to get everyone to agree on what counts as a conducive policy environment, a joint commitment, local ownership, local capacity, effective management, and effective monitoring. Without further specification, it is hard to see how these could be the basis for a framework which would be consensually agreed even within DFID, let alone with country partners.

1.6.2 A Skeleton for Frameworks

The Basic Axis of the Frameworks

There needs to be a framework of what to consider for each such group of performance indicators. The best starting point is probably the OECD/INES project suggestion of context, input, process and output. Given the focus here on projects and programmes, this should be extended to include aims and outcomes. It will, of course, not always be appropriate to consider each and every one of these components.

Stakeholders

In this context, who are likely to be affected by sets of performance indicators?

- DFID Education Advisors;
- programme participants;
- recipient governments;
- consultancy organisations;
- political authorities.

The concern to involve recipient governments (as well as direct beneficiaries) implies that, although there is no intention to develop a comprehensive framework, it is unlikely that a small set of 'key' indicators would be sufficient. The indicator set would have to reflect multiple goals with multiple indicators measured by multiple methods (McEwan and Hau Chow, 1991).

For example, an ODA survey of ELT projects (1994) identified 105 projects with an English focus or significant ELT component in 52 countries. DFID sees English within a wider development context of national language policy; indeed, the survey suggests that there is a distinction between English as a medium for educational development, for international communication, and for economic development. In contrast, the Department of Trade and Industry and the Foreign and Commonwealth Office, see English as an export. The survey concludes that at least three possible sets of indicators for English programmes/projects could be developed.

Relating Indicators to the Decision Making Context

Scheerens' distinction between types, levels and modes of decision-making is also important. What types of decision (choosing new projects, monitoring on-going projects)?; at what level (head office, country, project)?; in what mode (in-

house, with the country, for general public)?. The latter issue – of which 'mode' – is crucial in considering aid programmes and projects because of the range of 'stakeholders', but we should not forget the distinction between types of projects and levels of decision-making. The development of the skeleton frameworks suggested below for sets of performance indicators will obviously have to be appropriate to the specific task and the educational outcomes (cognitive/non-cognitive) considered. The corresponding indicators will vary depending upon the entry point (type, level or mode of decision-making).

Nevertheless, based on these principles, it should be possible to elaborate each of the lists in the previous section (one based on DFID strategy, the next on planning, the third on project implementation, and the last on economic appraisal) in collaboration with the appropriate personnel, field agencies etc. into an integrated and overlapping framework. The problem remains, however, that although an 'intuitive' reading of these different lists would find no contradiction, we have shown how the literature is replete with examples of what happens if one ignores the apparently small nuances between definitions of objectives and 'outcomes' when performance indicators are instantiated in the field.

Assuming these can be resolved, the crucial issue at each level is **what to include, what to omit and why.** Attention should be paid not only to the technical criteria for good indicators, but also to:

- issues of relevance to the particular system;
- propinquity to the phenomena being monitored without interfering in the operations of the system (either through an over-heavy burden of data collection or through giving the wrong incentives); and
- the potential need for multiple indicators because of multiple stakeholders.

1.6.3 Developing Indicators at Different Levels and for Different Stages

Essentially, for DFID, we can envisage three different sets of indicators: those at the sectoral level; those at the planning and pre-planning stage; and those which would be used for monitoring and evaluation.

Sectoral Level

If the focus is the sector, (and reflecting on the OECD experience) there is only a limited choice given the types of data that can be collected. The main problems are the mechanisms of collecting the data and its quality. Technology for carrying out sample surveys, however, is now well developed and, with country agreement, this is probably the best way forward.

Here the second axis of the table could be either the level or mode of decision-making. We have chosen to take the level as the crucial dimension, as this usually

determines the nature of possible participation. Here the entries in the cells are more in the nature of values relative to presumed targets for the sector (without falling into the trap of setting targets which will lead to falsification of data).

The table would have to be differentiated according to whether the focus of performance indicators was:

- the formal or informal system (CONTEXT);
- qualitative improvement or equity of access (AIMS);
- types of INPUTS and preferred PROCESSES;
- nature and timing of OUTPUTS and OUTCOMES.

This would, therefore, involve at least four tables according to whether the formal or non-formal system was being considered and whether the overall aim was qualitative improvement or equity of access. The variation between types of inputs and processes, and between kinds of outputs and outcomes (the latter closely connected with the AIMS) may not, however, be sufficient to generate substantially different tables.

Table 3: Planning Process and the Need for Performance Indicators at Different Levels

	Central	Regional	District	Village	Household
Context					
Aims					
Inputs					
Processes					
Outputs					
Outcomes					

For the non-formal system, the distinction made above in section 1.5 between different types of non-formal education and the corresponding types of performance indicators have to be taken into account as well as the distinctions made in Table 3. Lockheed and Levin (1991) suggest that context should be specified in terms of facilitating conditions such as community involvement, school-based professionalism, flexibility, and the will to act as reflected in vision and decentralised solutions. They do not specify exactly how one is meant to measure any of those and only seems feasible at the most local level. Their list of inputs, however, is probably as good a starting point as any: curriculum,

instructional materials, time for learning, teaching practices (Lockheed and Levin ,1991).

Planning and Pre-Planning of Projects

The approach here will probably have to be rather different. Consider the list cited in section 1.5.2.(2) above for the planning of projects:

What is the evidence of demand for the education service proposed? What will be the benefits to the country and individual? Can these improvements be measured and where possible quantified and qualified? Will there be cost economies resulting? Is the proposed strategy seen as the most cost effective? Are the recurrent cost implications manageable?

Recommendations: Each of these questions is likely to generate its own set of indicators - almost certainly too many for consistent judgements to be made - and each might well require a different set of data which would be costly to collect. Inasmuch as the answers to these questions are agreed to be important criteria for choosing projects, then perhaps the best approach is to propose what detailed specifications of indicators would have looked like in respect of a number of projects - some funded, some not funded - in order to assess whether or not those criteria are actually taken seriously, in deciding upon funding.

Project-Specific Performance Indicators

If we focus on the project, then it is appropriate to consider project-specific performance indicators (PSPIs). These are, in principle, straightforward measures of the extent to which a project completes the defined tasks: in the terminology of the logical framework, these are the 'purpose-level' indicators.

The issue is whether there is any scope for consistency in terms of the sets of performance indicators used in the logical frameworks for different types of projects (similar to the approach adopted by the Health and Population Division). This should be feasible for a large proportion of projects but the main problem is to assess exactly how these are used. At worst they may generate perverse incentives (as explained) or they may simply be ignored. In all cases, the importance of involving the 'beneficiaries' is crucial.

If we are considering a project then we need to be able to develop indicators for the context, aims, inputs, processes, outputs and outcomes (the same kind of list as in the logical framework) at different stages of the project cycle: distinguishing (at least) between pre-planning, start-up, mid-term evaluation and follow up. On this basis, an appropriate framework could be as follows:

Table 4: Performance Indicators at Different Stages of the Planning Process and of the Project Cycle

	Pre-planning	Start-up	Mid-term	Evaluation	Follow-up
Context					
Aims					
Inputs					
Processes					
Outputs					
Outcomes					

Essentially, the approach here is an extension of the project framework methodology which, although it is likely to miss crucial process characteristics, is used to assess the whole project process from start-up through to evaluation. That methodology was never intended to be used at the pre-planning phase; and although it is recommended for use as a basis for follow-up evaluation, the context may have changed dramatically, so that it may not be appropriate. Moreover, not all of the basic components would be appropriate at each stage.

In any practical application, of course, one would need to explode each of the cells, in terms of specifying the level at which indicators are required and, eventually, the extent of participation in the process.

ANNEX 1A

World Bank (1996) Performance Monitoring Indicators: A Handbook for Task Managers (Operational Policy Department)

The handbook specifies the potential uses of performance indicators for:

STRATEGIC PLANNING. For any program or activity, from a development project to a sales plan, incorporating performance measurement into the design forces greater consideration of the critical assumptions that underlie that program's relationship and causal paths. Thus performance indicators help clarify the objectives and logic of the programme.

PERFORMANCE ACCOUNTING. Performance indicators can help inform resource allocation decisions if they are used to direct resources to the most successful activities and thereby promote the most efficient use of resources.

FORECASTING AND EARLY WARNING DURING PROGRAM IMPLEMENTATION. Measuring progress against indicators may point toward future performance, providing feedback that can be used for planning, identifying areas needing improvement, and suggesting what can be done.

MEASURING PROGRAM RESULTS. Good performance indicators measure what a program has achieved relative to its objectives, not just what it has completed; thus they promote accountability.

PROGRAM MARKETING AND PUBLIC RELATIONS. Performance indicators can be used to demonstrate program results to satisfy an external audience. Performance data can be used to communicate the value of program or project to elected officials and the public.

BENCHMARKING. Performance indicators can generate data against which to measure other projects or programs. They also provide a way to improve programs by learning from successes, identifying good performers, and learning from their experience to improve the performance of others.

QUALITY MANAGEMENT. Performance indicators can be used to measure customer (beneficiary) satisfaction, and thereby assess whether and how the program is improving their lives.

The handbook recognises that the performance indicators must be based on the unique objectives of individual projects; but also that they should be based on an underlying logical framework that links project objectives with project components and respective inputs, activities and outputs at different stages.

They then discuss a number of advantages and limitation of Logical Framework and suggest a number of general principles for selecting indicators:

- relevance
- selectivity
- practicality of indicators, borrower ownership and data collection
- distinction between intermediate and leading indicators
- quantitative and qualitative indicators

There is then a description of the PMIs affecting the Bank's work at project identification, preparation, pre-appraisal, preparation/appraisal, implementation/ supervision, supervision/completion after completion onwards.

The following box illustrates the distinction between the performance information needs of differing levels of project management:

IMPLEMENTERS IN THE FIELD NEED

- input indicators
- output indicators
- [efficiency indicators]
- risk indicators
- some outcome and impact indicators

THE IMPLEMENTATION UNIT NEEDS

- summary input and output indicators, including site-comparative indicators as appropriate
- outcome indicators, including site-comparative indicators as appropriate
- [effectiveness indicators]
- risk indicators
- impact indicators

THE BORROWER AND THE BANK NEED

- summary input indicators
- summary output indicators
- risk indicators
- key outcome, impact [and relevance] indicators
- [sustainability indicators]

Note: Indicators in brackets are not a required part of Bank monitoring or project supervision.

But are they actually used in these ways?

ANNEX 1B

Problems of Measurement at the Sectoral Level: Examples of Indicators and their Associated Problems

The intention of this section is not to provide an exhaustive overview of what needs to be measured at the sectoral level. We have already explained that the detail of performance indicators has to be developed in conjunction with in-country representatives. Instead, the purpose is simply to draw attention to some problems of definition and data where the performance indicators are to be based on data from the entire system.[8]

Enrolment Ratios

The problem here is 'simple': the quality of the data. There are several aspects to both the numerator and denominator:

- what is actually meant by enrolment: registration of child (for whatever reason), appearance in class at beginning of school year (if that can be identified), regular (rather than sporadic) attendance, inscription for (or sitting) annual examination;

- the relevant population: the 'decennial' censuses in many developing countries are unreliable for a variety of reasons so that the estimated size of the relevant age group has to be treated with extreme caution (Murray, 1988).

Attendance

No one suggests that this will be easy to monitor in developing countries. Yet, after enrolment, it is the next most important statistic because, without that data, we cannot sensibly assess what the enrolment figures mean in terms of childhood exposure to school. It might give some perverse comfort to know that this has also been a problem in developed countries. Ruby (1992) explains the difficulties of operationalising attendance as part of the OECD project set of indicators.

Measuring Quality

This is probably the most contentious area. Most of the heat has been generated around school effectiveness research because of the growth of achievement monitoring in order to identify "improved environments and educational aids which lead to detectable gains in knowledge, skills and values acquired by students" (Ross and Mahlck 1989).

---

[8] Many of the problems described in this section can be overcome when a statistically representative sample is taken, because much more effort can be put into securing the quality of the data. The move towards decentralisation, however, implies collecting data on all those in a smaller unit.

In addition there are arguments over what is meant by 'quality' (e.g. Cheng, 1994) and over who should decide what is meant by quality (e.g. Hoppers, 1994; Stephens, 1991). Together, this would suggest that it is foolhardy to propose a system intended to be valid across all countries and systems

Instead, it might be appropriate to consider the more cautious approach of the USAID research project on Improving Educational Quality (IEQ) (with offices in Mali, Ghana, Guatemala, South Africa and Uganda) with the following objectives:

- to understand the processes through which classroom interventions in different countries influence student performance;
- to demonstrate a process whereby classroom research on improving educational quality is integrated into the educational system;
- to create opportunities for dialogue and partnership among researchers and educators who are seeking to improve educational quality at local, regional, national and international levels.

In this way, the intention, presumably, is to develop sets of quality indicators that are consensually agreed at the country level. Whether this is feasible, and whether the conclusions of such groups actually do generate national consensus is unclear.

Disadvantaged groups

Both the Jomtien agenda and the ODA Education Strategy Paper (1993) highlight the importance of monitoring the situation of disadvantaged groups. Possible indicators are:

- participation and success rate of ethnic, religious or language minority students;
- number and status of teachers and administrators in the educational system from those groups;
- appropriate curriculum and textbook content;
- provision of teachers familiar with non-mainstream cultures; and
- linguistic information on teachers and students.

This is very obviously a case where no hard and fast suggestions can be made. It depends on the specific situations in each country or region. Introducing a term or concept from one country or region to another may lead to entirely inappropriate conclusions.

What Happens After School?

The ODA Education Strategy Paper (1993) points to the difficulties here of institutionalising verification and accountability mechanisms for the wider objectives and longer term outputs of projects[9]:

- on exit from education system, 'no one is responsible';
- the management and design of tracer studies is rarely specified clearly in project design;
- interest in longer term outputs/outcomes often diminishes.

From the societal point of view, this is the most important outcome; yet the 1999 Policy Framework Document is much vaguer, talking about strengthening capacity (p40) and rights and responsibilities (p33).

Decentralisation and Devolution

Indicators of the devolution of financial responsibility can include: number of distinct school systems; proportion of key education decisions that are made locally; existence of school boards, their methods of selection and financial mandates; percent of locally generated revenue that stays local.

In the health sector of Scandinavian countries, Mills suggests assessing the extent of decentralisation in terms of the following two sets of indicators:

( ) revenue raising in devolved systems (thus: percentage of public health care centrally funded; local authority tax powers; controls on local taxes; central sanctions if expenditure is exceeded; and the local right to take out loans);

( ) planning controls in devolved systems (thus: the existence of a planning process linking levels; the initiating level; whether it is compulsory; and whether government approval is required.

Democratisation/Beneficiary Participation

The World Bank (1996) now argues for monitoring beneficiary participation to increase client investment in project success. Developing joint monitoring and evaluation systems work towards Bank goals of teaching new skills, but one must

[9] Indeed, according to an DFID Internal Document (December 1995, para. 23) "Experience suggests that it is not worth spending too much time trying to identify indicators at the level of wider objectives, as these are unlikely to be very relevant to the project itself. If necessary, the wider objectives box for indicators can be left blank!" (author's exclamation mark).

note their own caveat that these require: continuity of personnel from both government and donor agency; a network of supportive government personnel; avoidance of partisan politics; community leadership; and a sense of community and investment in project goals (Uphoff, 1992). Clearly, not an easy task.

ANNEX 1C

Collecting Data for Individual Performance Indicators

It is not clear exactly what should count as the final outcomes for individual pupils or for a group of pupils. As we have already emphasised several times, it depends on the original objectives of education in the first place. Schematically one could distinguish between those who emphasise individual (educational) attainment, others the kind of job/income and therefore opportunity for social mobility for people, and yet others the quality of life that people lead.

Conceptual Problems

The appropriate indicators of outcome or performance would, of course, be different in each case.

- income/jobs/social mobility In principle, the indicators are based on the relation between the years of schooling and estimated lifetime earnings or the different jobs statuses. Putting to one side for the moment the well-known difficulties of collecting the data (see below) there is also the problem that such analyses assume there is an income or jobs to go to, and also that we can count the amount of education in terms of number of years of schooling. The former problem is one of many concerned with the problem of interpreting such analyses which have been dealt with thoroughly elsewhere (e.g. Hough 1990); but the latter has often been ignored (or assumed away which amounts to the same thing). Yet we all know that a year of schooling in Sweden is something very different from a year of schooling in Zambia; although we can crudely account for that by taking the cost of a year of schooling as the measure of resource input (rather than counting years), and the same procedure could be followed for non-formal education programmes. But, *within a system,* the quality of a year of schooling can vary enormously with only minor - or no discernible - variation in resource input and there is, no obviously way of adjusting for this apart from using another outcome measure (such as attainment levels) which means we would not be able to calculate any ratio of labour market outcome to the resource costs of the effort required. Basically, it means that data has to be collected on quality as well as quantity of schooling.

- quality of life : Although one has the same problem of counting years of schooling, the corresponding difficulty of the appropriate measure of outcomes is rather different and less attention has been paid to it in the literature. Part of the answer is a systematic attempt to measure the quality of life as illustrated in Part III.

- individual attainment The indicators appear 'obvious' here, although the problems of interpretation are often underestimated. Even assuming agreement on the range of curriculum topics/subjects which should be the subject of measurement, the difficulty is in separating out the background and school influences (the value added problem). This is, of course, a contentious issue in developing countries; but in some respects, separating out home and school effects is a simpler problem because the background of pupils/students is so similar. On the other hand, there is very little data.

Collecting the Data

In addition to these 'conceptual' problems, there are also considerable difficulties with collecting the data in developing countries. There are two main choices: following a cohort of pupils/students through from school to the labour market (what are sometimes called 'tracer' studies); or collecting retrospective data on a random sample of adults in the labour market (what could be called 'reverse tracer' studies). Each pose different sets of problems.

- tracer studies. The obvious problems here are the difficulty and expense of following up people for any length of time: keeping track of people's movements is complicated. However, this has the obvious advantage of collecting data at the time it happens (whether one is talking about the quality and quantity of schooling OR about the income/job).

- reverse tracer studies. These studies are, in principle, much cheaper and easier to organise, as they simply involve a questionnaire to adults about their school experiences. But, inasmuch as one believes that the quality of schooling is an important variable, it is obviously inadequate to rely on adult recall of their schooling experiences. This therefore entails identifying the schools that the adults attended and then retrieving the files on them (if they exist and can be found). This exercise is also tedious and time consuming, although not usually as much as the prospective approach. However, it is unlikely to be 100% successful because the files might have been mislaid or erratically filed or simply because they are incomplete. One cannot therefore rely on this procedure to provide good data on quality.

CHAPTER TWO

CASE STUDIES

CHAPTER TWO

CASE STUDIES

2.1 Introduction

In the preceding chapter, a number of propositions based on the literature have been developed together with a proposed framework which could eventually be used at the sectoral level at different stages. The purpose of this section of the report is to examine the extent to which these theoretical notions might apply in a number of different contexts.

2.1.1 The Choice of Case Studies

Three very different contexts have been chosen. In Kenya, independent in 1960, there has been a long-standing ODA/DFID involvement in a variety of education projects ranging from support to self-help school-building to encouraging the development of an indigenous publishing industry. Whilst there has been a general drift towards improving educational management and quality improvement via providing local support to teachers, and towards consolidation within one overall programme, SPRED (Strengthening Primary Education) Phases I and II; there are still clearly distinct components.

In Andhra Pradesh (an average sized Indian state with a population of approximately 60 million), there has been a large scale, state-wide, unified programme since 1984. The programme emphasises combining inputs at the school level and has been implemented using a rolling approach across the state.

Finally, the case of South Africa is seen as being of particular interest because it is newly independent and because of the problems of integrating a large number of pre-existing educational systems into a coherent whole.

2.1.2 The Approaches to Evaluation

In Kenya, evaluation has always been highly contextual. For example, the review of the informal sector by King (1996) was mainly an historical analysis. The evaluation methodologies promoted under SPRED I were highly qualitative without being at all rigorous; and so on. Moreover, the evaluation with SPRED I had been university based even though there had been difficulties in the university during this period.

In Andhra Pradesh, the approach to evaluation has been entirely different with the appointment of external consultants to carry out an ongoing quantitative evaluation. This led to the development of a highly quantitative, formal evaluation scheme with schools being rated on their degree of APPEPness (see Cooper *et al*, 1996) which was seen as demonstrating the success of the programme, and the extent to

which they had absorbed the principal features of the Andhra Pradesh Primary Education Programme). At the same time, it has appeared to generate some internal criticism, possibly through lack of local ownership.

Finally, although there had been a functioning system of education statistics and indicators in South Africa, the form and content of that system was clearly only relevant for the preceding regime. The purpose of the case study here was to examine what kind of performance indicators were being developed in the new situation. Given the embryonic state of development, however, the discussion in section 4 of the situation in South Africa is mostly concerned with what might the implications of developing structures and the extent to which key actors in the system are aware of the pitfalls of different kinds of data and statistics.

2.2 MEASURING THE IMPACT OF EDUCATION IN KENYA

2.2.1. Introduction

The purpose of this section is to examine, briefly, the sorts of indicators being used by the authorities in Kenya to judge the impact of its educational policies. The chapter starts with a short review of the education system in Kenya, but the reader is referred to a DFID (formerly ODA) publication by Oketch (1995) for a fuller description and review.

2.2.2. Schooling in Kenya

Education is not free in Kenya and this leads to considerable hardships, particularly for parents in the poorest categories. This category can reach quite far. For instance, unskilled workers in urban areas earn around Ksh2000 to 3000 per month in 1996 (Ksh80 = £1) and rural workers make do on around Ksh70-80 per day that is when there is work. Costs of sending a child to school vary from Ksh300 to 700 a month. This makes it difficult if not impossible for poor families with four or five children (the norm with population growth running at a high 2.9% p.a.) to send their children to school. And, if a decision is taken to send one child to school, it will usually be male.

Many items (buildings, school meals, books, transport, etc.) are paid for by the parents themselves. Even for handicapped and special needs schools the main financial burden must be carried by the parents. Some contributions are received from outside donors or other institutions for special schools or for schools in the poorest areas, such as the arid regions of Kenya, and for mission schools. There are about 16,000 primary schools in Kenya, 3,000 secondary schools, 200 diploma-level colleges and 6 public universities.

The Ministry of Education (MoE) in Kenya concentrates on overall policy development, monitoring of the educational system, planning future expansion based on projected numbers enrolled, training teachers, monitoring examinations and curriculum development. The Ministry of Technical Training co-ordinates vocational and technical training in the country and observes the technical

standards and certification set by the Kenyan National Examination Council (KNEC).

The Ministry of Planning and Development also uses data from the MoE to prepare its plans - the most recent development plan covers 1997-2001(Government of Kenya, 1996a). Currently, too, the MoE is preparing a long-term educational plan over 1997 to 2010. The overall plan focuses upon industrialisation and the objective is to move the economy to the status of an industrialising (not industrialised, note) economy by the year 2020. This will require raising annual savings from the current 17% of GDP to 30%. The Plan does not ignore unemployment and poverty and states that the long term objective of the Government is to reduce these, with youth and other disadvantaged members receiving special attention. It does ignore education: the Plan re-states the Government's guiding philosophy, that every Kenyan has the inalienable right, no matter his or her socio-economic status, to basic education. Yet, the 1997-2001 National Development Plan does not give a high priority to education, particularly to the primary education that is so sorely required.

The main problems of education in Kenya are low quality of instruction, uneven standards, high repetition and dropout rates, especially for girls, declining enrollment rates, and serious inequities in the system. For instance, pupil-teacher ratios in primary education vary from 1:10 to 1:70 (World Bank 1997a); and funds are disproportionately allocated - current public expenditure is 57% primary, 16.2% secondary and 20% tertiary while enrolments are 89%, 29% and 2%, respectively. The richest 10% of schools receive more than four times the subsidy from the public purse than the poorest 10%. Due to structural adjustment policies, there has been a reduction in pubic expenditure at a time when cost-sharing by parents has probably peaked (parents simply cannot give any more), and this has created a resource gap that is likely to lead to reductions in educational quality and enrolments (Government of Kenya, 1996b). This can already be seen with the rate of growth of enrolment of 6-14 year-olds being less than 2.5% while the population in that age range is growing at about 3%. For primary education, over 1989 to 1995 enrolments fell from 95% to 79%.

The financing of education has continued to trouble the Government. The Central Government makes substantial inputs to education. Government expenditure doubled over 1992 to 1995, and the expenditure of the Ministries of Education and Research, Technical Training and Technology has been increasing steadily. Currently public expenditure on education is 6.2% of GDP and 38% of total public recurrent expenditure (Abagi, 1997). The Government's Plan states that the policy of cost-sharing between the Government, parents, and communities, which has been in place since 1988, will be maintained. It notes that the bulk of Government subsidies are in the form of teachers' salaries which have led to "problems for poor students who cannot afford to pay for books or equipment" although it will take "measures to rectify these adverse trends" without specifying what they mean. The fact that teachers' salaries are fixed and not based upon delivery has not been addressed. Nor does it take account of the serious problems in the arid and semi-arid areas of the country where education is in competition with survival needs. But the main problem is inefficiency in the allocation of public expenditure to education.

Despite relatively large sums being spent on education, quality is declining and enrolment levels falling.

The ODA financed study by Oketch (1995) is pessimistic that the six major reviews of educational policy in Kenya since independence will be successful. These reviews have suggested alternative models of education and training that culminated in changes in the mid-1980s, introducing, for instance, the 8:4:4 system of education. Oketch notes that whilst "the structure and content of the curriculum has been changed to emphasise the learning of practical skills, the strategy is unfortunately based on an inadequate assessment of resources and other important aspects of education such as the ability of the students to learn certain concepts and skills at certain ages or times. The current system was introduced in the absence of teachers with experience and skills in teaching the newly introduced practical skills. The government also ignored or underestimated the cost of financing the teaching of practical subjects in schools."

2.2.3. VET in Kenya

Kenya's vocational training and technical education system (VET) suffers from some of the same weaknesses that characterise many similar systems in developing countries that have been fostered and financed entirely by the public sector (Kilele and Sinclair, 1991). That is they tend to be orientated toward modern sector employment, provide inflexible training courses with rigid curricula and forms of examination which bear no or very little relation to market demand, and which operate under fixed and inflexible budgets with little to no discretion available to directors of such centres.

There is a bigger demand for VET school places than capacity at the same time as there is a growing problem of educated unemployed; hence a mismatch between skills supplied on the labour market from VET and its demand. Yet VET only accounts for about 3% of public expenditure on secondary education. The system is skewed, too, in favour of science subjects rather than more practical skills required by industry (this is particularly important given that the next national plan proposes a major thrust into industrialisation). This has resulted, for instance, in a current shortage of engineers and a surplus of science graduates.

Observers believe that the imbalance between jobs and the labour force in Kenya will worsen, not improve in the future (Oketch 1995). The Eighth Plan has recognised some of these difficulties and notes that on-the-job training will be emphasised in both public and private sectors given "its superiority to training in formal institutions". A review of the requirements for scientific manpower is planned for the first year of the Plan and the Government intends that its public training institutions will become more "demand driven" with client-centred training to meet the needs of the service and thereby be able to compete effectively with the private sector.

However the information base with which to do this is woefully weak. As the Eighth Plan notes, "there is inadequate data for manpower planning...there is no

comprehensive system for monitoring human resource trends...there is no validation mechanism for comparing actual training performance with targets...the occupational classification in Kenyan industry is haphazard and does not adequately take account of changing skills and technology."

2.2.4. Kenyan Data

In an ODA document produced under the SPRED I project in 1995 (ODA 1995), it was noted that:

> The MoE produces vast amounts of data at school level. However its capacity to retrieve and analyse this for the purpose of producing reports on performance, economic planning and improving efficiency is extremely limited. Under the project the MoE has provided a full staff establishment of professional officers in the PU [Planning Unit] to allow analysis of access to education. ODA has assisted with physical resources (computers), staff training (in-country) and specialist expertise for the development and initial operation of an institutional database.

The situation has changed little at the time of writing.

National Level Data

In 1993 and 1995, under SPRED I, the Ministry of Education circulated a questionnaire on statistics of primary schools and another one on the statistics of secondary schools. All primary and secondary schools (both public and private) were required to complete the questionnaire. This was in addition to the normal educational census data collected annually (see below). The questionnaire was sent out to assess enrollment, dropout, repetition rates, teachers' qualifications, but also to assess the reasons for dropout or non-enrolment in primary education. Response rates of up to 95% were obtained, yet at the time of writing (March 1998) published results were not available for the 1993 survey, although printouts could be obtained from the MoE on request; and data were still being keyed into a computer from the 1995 returns. Educational census data for 1995 were available, however. Listed below are the categories of data collected for primary and secondary schools.

Table 5:Annual Educational Census Data

School particulars (name, address, telephone, name of sponsor, province, district, division, zone), year and month established, date of school's last inspection)
School type, status, attendance (public/private; mixed, girls only, boys only; day/boarding/day&boarding, if day&boarding number of pupils)
Enrolment data (by class, stream, gender and age)
Repeaters (by class, sex)
School milk receivers? (yes/no, numbers by UHT/pasteurised)
School Feeding Programme (yes/no, Programme type by sex)
Data on Special Education (school category, type of disability by sex by school category)
Teachers (by qualification and gender, age, teaching experience)
Teachers by employer (type by sex)
Teachers leaving school (by reason, trained/untrained, sex)
Examination results (subject, sex, grade in Kenyan Certificate of Primary Education (KCPE), level in Kenyan Certificate of Secondary Education (KCSE) exam)
Physical facilities & School Equipment (land size by use/buildings by type, number completed, number under construction, number planned/school equipment by classroom or office, by type, total required, currently available, shortfall/ accommodation type)
Other facilities (water supply/mode of lighting/mode of cooking/access to health facilities)
Grants/Aid (yes/no, by type, amount)
School fees paid by parents per student per year (by use, by class)
Dropouts and transfers (by sex and class)
Reason for dropout (reason by sex)
Non-enrolments (reason for non-enrolment by sex)

Other survey data have been collected as well:

- under the PRISM (Primary School Management) project a questionnaire was sent out in 1995 to all headmasters to assess the quality of in-service training;
- a welfare survey, at the household level, was carried out in 1994 and in 1997 (World Bank 1997b)[10]; and
- tracer studies were carried out in the early 1980s to measure the impact of VET.[11]

In addition, DFID has installed a management information system in the polytechnics and these may contain some placement data.

Education and Employment Data

Table 6 summarises the data available to assess the impact of education on employment- related indicators. (See Chapter 3, section 3.7).

How are education data used by national authorities?

The Ministry of Education uses these data to plan for the level of teacher training required over the coming years, based on projections of school enrollments less dropout and repetition rates. It is also concerned about dropout rates and uses these data to hone its policies. To this end it is currently carrying out a medium-term plan to assess future resource needs. Results are also published in the Government's annual *Economic Survey*, which has a chapter on education and in the annual *Statistical Abstract* published by the Central Bureau of Statistics.

[10] The WMES (Welfare Monitoring and Evaluation Survey) includes some questions on the educational level of individuals within the household. The surveys, following typical World Bank LSMS (Living Standards Measurement Survey) procedures, covered a sample of around 10,000 households nationally.

[11] It seems that since then, nothing has been done. Apparently the training schools are oversubscribed but the economy cannot absorb them all. There is clearly a case to assess the rate of return of VET vs. general secondary education (see Bennell 1996) to see what type of courses are being offered and to see why enterprises are not absorbing all the output.

Table 6: Education and Employment Indicator Availability

	TYPE OF EDUCATION			
Suggested Types of Indicators	**Primary**	**Secondary**	**VET**	**Tertiary**
Earnings (wages, other earnings)	1987 Manpower Survey?	1987 Manpower Survey?		1987 Manpower Survey?
Rate of return (private and social)	?	?		?
Employment status (not in labour force, unemployed, underemployed, employed)	1991 Labour Survey?	1991 Labour Survey?		1991 Labour Survey?
Poverty status (less than food-based poverty, ultra-poor)	1992 Welfare Monitoring and Evaluation Survey	1992 Welfare Monitoring and Evaluation Survey		1992 Welfare Monitoring and Evaluation Survey
Mis-match between job obtained and job for which qualified		1995 Assessment of Scientific Manpower Req'ments		1995 Assessment of Scientific Manpower Req'ments
Responses from employers on appropriateness of training received				Not systematic
Satisfaction gained from work	?	?	?	?

An interview at the Ministry of Finance to investigate whether education data are used to prepare the budget showed that this was not the case. Instead, as is done in many countries, the current year's budget is based merely upon the previous year's, using a multiplier, and added to which is an adjustment made for inflation. The Ministry of Education is then given a ceiling within which it must keep. Changes to this are possible in theory, should the Ministry of Education "make a strong case to get more money" but this seems rare in practice.

Plans for improving quality and range of data?

The MoE is aware of the weaknesses in its statistical base. A major problem is that vast data sets are collected but the data are so raw that they need significant processing. This is slow due to lack of motivation of staff and the sheer volume of data. For instance, the primary and secondary school questionnaire in 1995 covered 20 pages and contained about 150-200 data variables. The quality of these data is not as good as it should be because school respondents have no training in filling in the forms, nor do they regard them as a high priority and, in some cases, enrollment numbers are exaggerated to attract more public funds. The same is true of the annual educational census. The MoE would like more funds to train data collectors, to do more than just have telephone contact with District Education Officers (DEOs) and to set up on a regular basis a well organised MIS for educational data. The MoE is also hoping that the World Bank, in a future educational loan, will help them improve their data base. Expenditure data, apparently, are also in disarray. "No established mechanisms exist to capture the entire expenditure in education." (Government of Kenya 1996b)

Data in the VET area are in a worse state. There is no statistical section in the Ministry concerned, despite their needing a proper information system. Nor did it seem that plans were being made to improve the situation.

Plans for using indicators at a sub-national level?

It appears that little use is made of educational statistics to organise the education sector, budgeting is done on a "more of the same" basis. Consequently, sub-national planning is in a worse shape than at the national level and there seemed to be no plans designed to address the situation. In fact Kenya is the only country in East Africa that does not publish education statistics.

2.2.5. DFID in Kenya

With this background, DFID has supported primary education with about £4mn under SPRED I (Strengthening of Primary Education Phase One) over 1992-96, and SPRED II (Strengthening of Primary Education Phase Two) has allocated £18.7mn from 1997-2001 (GoK £16.69mn). The overall goal of SPRED II is to increase demand for and the utilisation of high quality primary education and it will focus on educational outcomes, institutional reform, financial reform, and empowering parents. This focus emerges, in particular, from the fact that over the period of SPRED I primary school enrolments actually declined. An additional amount of £4.6mn has also been allocated to upgrade 15,000 head teachers through in-service training (the Primary School Management (PRISM) project). DFID has also assisted with the strengthening of the in-service teacher development system through a network of teacher advisory centres (TACs).

A SPRED I evaluation in early 1996 (ODA, 1996) highlighted significant achievements in strengthening the teacher in-service and schools advisory support system. Nevertheless, the evaluation pointed to a need to shift the balance of both programme planning and resources towards school/community levels in

order to impact more strongly on the 'achievement culture' of the school. The evaluation proposed a greater focus on TACs as a community education centre, including incentives and mechanisms for greater community participation in the planning and management of teacher development, targeting girls and health/nutrition programmes.

Uses of Indicators in Programme Evaluation

The evaluation report also noted that there was neither a well-defined monitoring programme, nor were the monitoring responsibilities of various parties made clear. Further, the absence of a comprehensive baseline survey was a constant limitation for monitoring progress and impact.

The main indicators identified to measure progress with respect to the SPRED II goal to "increase demand for and utilisation of high quality primary education" are fourfold:

- *reduced wastage rates*, especially for girls, from the current 56% to less than 30% by 2005;
- *improved student performance*, raising the average achieved on the Kenyan Certificate of Primary Education (KCPE) by 20 points; and stabilising repetition rates at 15% through to 2005;
- *increased gross enrollment rates* (GER) from the current 79% to 85% by 2005; and
- *increased pupil and parental satisfaction.*

The logical framework also includes a list of 45 further, so-called "measurable indicators" to measure progress on purpose, outputs and activities. Most of them essentially measure inputs. For example, to improve teacher training through the School-Based Teacher Development Programme (STD) would mean, *inter alia*, testing and measurement booklets prepared for circulation to 175,000 teachers by KCPE examiners by end of 1997.

The fourth main indicator (pupil and parental satisfaction) is difficult to quantify and measures are not proposed in the logical framework. More precise indicators on the "demand" for education would be useful; and one can note that no links to the labour market are proposed in any of the 4+45 indicators.

A Learning Process

According to the evaluators:

"the performance indicators and monitoring mechanisms identified for SPRED are unhelpful ... at the top two levels the indicators are given in vague qualitative terms, even for essentially quantitative data. So, for example, an indicator ... of 'improved equitable access' is '4.3 Reduction in Gender Differences' [which] may involve gender disaggregation of enrolment statistics, KCPE results, retention or repetition rates, or a qualitative measure such as change in the way teachers interact with pupils of different sexes. Quantified targets for specific measures must be the goal if project description is to be the basis of agreement and implementation activities"(ODA, 1996:11).

While recognising the difficulty of finding such specific indicators where the objectives are essentially qualitative, the evaluators suggest that more effort should have been made. For example, for the intermediate objective 'institutionalisating the primary level TACs,' the indicators are:

a) TACs working effectively

b) Use of (TAC tutors and subject) handbooks

c) Increased administrative support

The evaluators comment that the effectiveness of TACs might have been indicated through records of courses and visits; standardised post-training evaluations by teachers; and record of utilisation by teachers of TAC resources.

They conclude that: "We could not find any evidence that incremental improvement in teaching quality had helped improve enrolment or reduce wastage and repetition. Neither did we identify any parents for whom additional financial demands resulting from SPRED I could be identified as having been the 'last straw'". (ODA, 1996:36).

However, this appears to be a general lament: indeed, it is curious to note that, despite the evaluator's emphasis on the development of specified indicators, in their own commentary on the impact of sustainability of SPRED they make no reference to the highly quantified and specific targets in SPRED II (focusing on wastage rates, the KCPE average and retention rates, the GER and satisfaction). This is possibly due in part to the difficulties of collecting the appropriate data.

Moreover, the 'lessons' of SPRED 1do not appear to have been learned for SPRED II. The indicators are still as vague as they were in SPRED I:

Purpose 1.1 Teaching and learning environment improved in all districts by 1999, through all teachers using new skills that inspire active learning and through use of textbooks provided under project.

Purpose 1.2 Improved professional support and inspection services to schools nationwide through ungraded and diversified Teachers Advisory Centre (TAC) system and upgraded inspectorate by 2000.

[Taken from their Annex D: this is an early log frame for SPRED II]

2.2.6. Concluding remarks

In Chapter One, a "generic framework" for defining indicators was presented based on Oakes (1986). Table 7 reviews this framework as applied to the current situation in Kenya.

In conclusion education is well-funded in Kenya but its impact is poor and declining. It would be useful for the Government to use a few benchmark indicators to measure progress of what they actually get out of the approximately 30% of public recurrent expenditure they spend on education. There seems to be no understanding of this at present

Table 7:Generic Indicator Framework Applied to Kenya

Indicators must provide at least one of the following kinds of information:	*And in Kenya?*
* a description of performance in achieving desired educational conditions and outcomes	* the Plan attempts to do this but data are weak
* features known through research to be linked with desired outcomes	* maybe the medium term plan will have some of this but none in evidence to date
* a description of central features of the system in order to understand its functioning	* this seems to be available
* problem orientated	* this doesn't appear to be the case
* policy-relevant	* not systematically so especially for VET
Indicators should have the following technical characteristics:	
* ubiquitous features of schooling found in some form throughout the systems/settings being compared	* this is done but problems in organising data at national level
* enduring features of the systems so that trends over time can be analysed	* done to a certain extent for basic indicators
* feasible in terms of time, cost and expertise	* expertise is available; resources are the problem
* generally accepted as valid and reliable statistics	* this does not seem to be the case

2.3 PERFORMANCE INDICATORS IN THE ANDHRA PRADESH DISTRICT PRIMARY EDUCATION PROJECT[1]

This section is rather different from the others: it reports on the design of performance indicators for the evaluation of a specific project, the Andhra Pradesh District Primary Education Project (APDPEP). First of all, it is important to understand the overall context of the project and then of the design of the evaluation which was intended to generate the maximum possible participation among project beneficiaries. In particular, the evaluation of the previous project in the same state (the Andhra Pradesh Primary Education Project (APPEP) was seen as crucial for both the project (hence the current district focus) and for the evaluation.

2.3.1 The Development of the Project from its Precursors

The overall goals of the projects have been to increase and extend the levels of literacy and numeracy in Andhra Pradesh, and the immediate purpose is to improve access especially to disadvantaged groups to quality primary education services (formal and non-formal) in five districts. During the original Andhra Pradesh Primary Education Project (APPEP), a rather classical approach to evaluation was adopted based on large-scale surveys of outcomes measured by educational attainment tests.

There had been proposals to introduce more process and 'qualitative' types of measure (Rao 1997). Some of those suggested were, for example:

Delivery of Inputs

- Provision of APPEP initial in-service training
- Numbers of visits made by Mandal Education officer
- Support of Colleagues as reported by teachers.

Effectiveness of Inputs

- Reaction to in-service training
- Percentage of teachers participating in different activities of Teacher Centre
- Percentage of teachers regularly conducting group work
- Reaction of teachers to numbers of teacher centre meetings
- Pedagogical Group Work and Questioning and Pupil involvement

Outcomes of Impact

- Numbers of visits of parents to schools

---

1 Revised and updated version of this section has been published in DFID India (1998) Evaluation in Primary Education – a Handbook for getting Started, New Delhi, and in McKay and Treffgarne (1999, eds.), Assessing Impact, London, DFID. The focus here, however, is on the development of the design rather than on any specific performance indicators.

- Reported change in children's behaviours
- Children kept away from school
- Pupil enjoyment of school (pupil response to APPEP)
- Decrease in absenteeism
- Enrolment
- Attendance
- Routine test scores
- 'Assessment'

The APPEP evaluation ran into difficulties of sustaining local capacity in the absence of an institutional structure. As a result, there was no formative evaluation (of the kind described above) and the feedback from the summative evaluations was slow ; and the impacts were studied only at the school level, ignoring both the individual and community level. Basically, the problem was that, whilst all sorts of interesting data were being produced from the study, these were only useful as performance indicators in a top-down mode: what was and is required were indicators which would be immediately useful for project management – with the twist that the project management was itself (as far as was possible) participatory.

The problem, therefore for the current (APDPEP) project, was to design a system of performance indicators, or mechanisms for collecting data with which to construct trial performance indicators, which would reflect this participatory approach to project management; and which would help build an informed network of community 'experts'. In this situation, the usual distinction between monitoring (a continuous management activity providing regular feedback to those hierarchically responsible for implementing the activity) and evaluation (an external means of assessing programme results and the appropriateness of the design in achieving hierarchically specified objectives) is unclear. Communities are continuously assessing process and outcome and, *where problems are not due to **absolute** lack of resources*, it is the communities that are best placed to examine those problems and identify solutions.

These considerations led to a three-stage/phase evaluation design: a school and pupil survey; a suite of qualitative studies; and a set of Indicators of Implementation.

2.3.2 Quantitative and Qualitative Studies: Project Planning Process

School and pupil surveys, repeated over the lifetime of the project, were to measure the cumulative impact of the project against the principal project objectives and planned outputs. These would establish what the effect of the project has been and which children from different communities and socio-economic groups have benefited. The results, of the survey would portray the statistical linkages between enrolment, retention, parental literacy, economic status, drop-out and learning achievement.

In order to understand the processes, it would then be important to carry out qualitative studies, to provide insight into the patterns of success and failure

uncovered by qualitative surveys. The combination provides project mangers at various levels with an understanding of the impact of the project. (See Figure 4).

2.3.3 **Qualitative Case Studies**

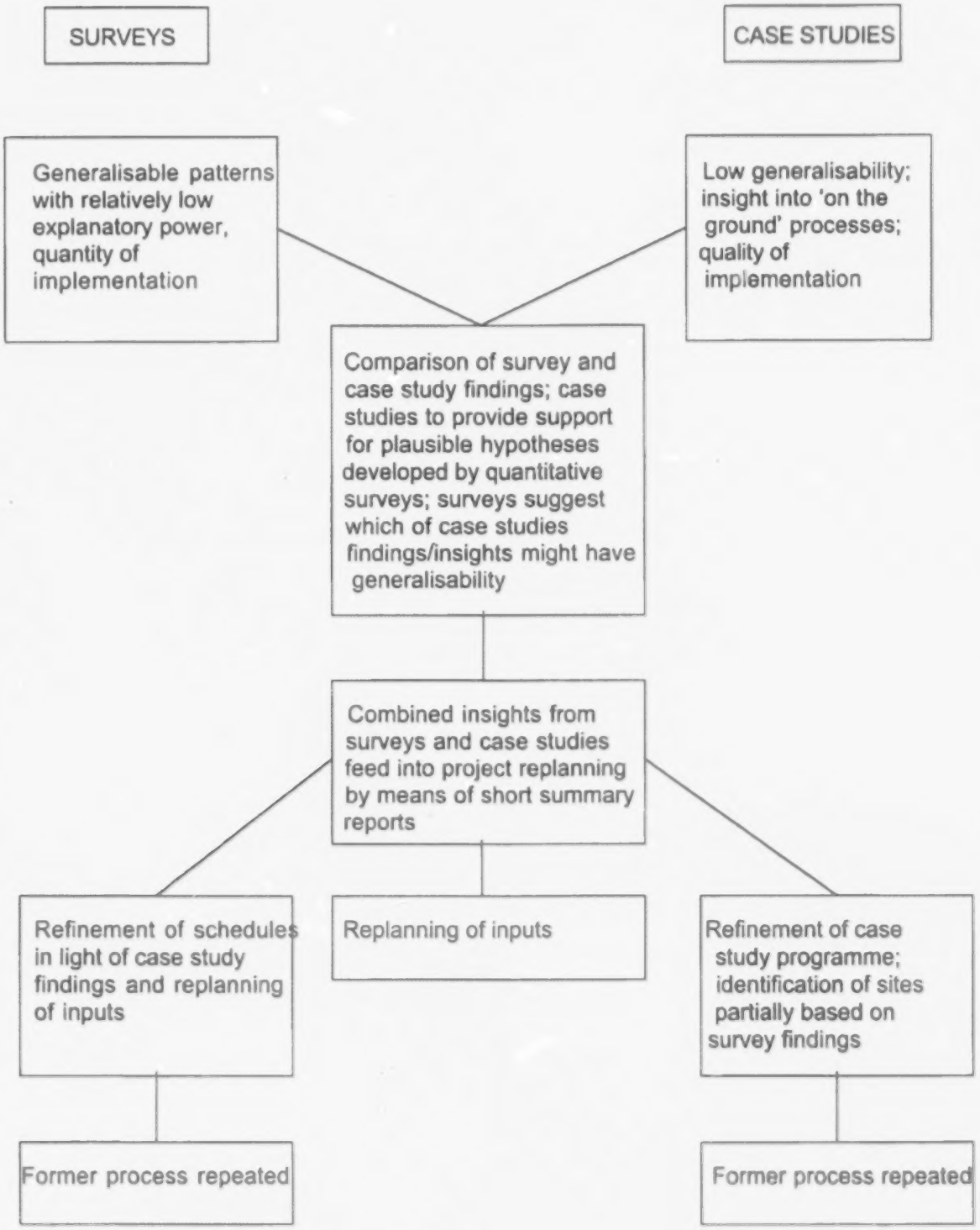
In addition to complement the findings from the quantitative surveys, a variety of other kinds of case studies were proposed.

- one-off case studies of a 'good' school in different districts or the constraints experienced by particular schools in retaining girls;

- linked and/or longitudinal studies of change and development among schools and development among schools and Village Education Committees (VECs).

The qualitative case studies serve to complement the others in two ways. First, while the indicators of implementation provide rapid quantitative feedback, the case studies provide qualitative feedback speaking directly to teachers and community members through vivid and accessible accounts of strategies and practices in different classroom settings. Second, the case studies complement and extend the survey analyses by providing insight into, and explanation of particular patterns of progress (among a purposive sample of pupils, schools or local support agencies).

Figure 4: Quantitative and Qualitative Studies: Project Planning Process


SURVEYS
CASE STUDIES
Generalisable patterns with relatively low explanatory power, quantity of implementation
Low generalisability; insight into 'on the ground' processes; quality of implementation
Comparison of survey and case study findings; case studies to provide support for plausible hypotheses developed by quantitative surveys; surveys suggest which of case studies findings/insights might have generalisability
Combined insights from surveys and case studies feed into project replanning by means of short summary reports
Refinement of schedules in light of case study findings and replanning of inputs
Replanning of inputs
Refinement of case study programme; identification of sites partially based on survey findings
Former process repeated
Former process repeated

2.3.4. Indicators of Implementation

An attempt was made to analyse the project concerns with planning and implementation in terms of five broad areas for action and to ask, in each case, three further questions. First, what processes would need to be set in place to ensure that this priority is reached? Secondly, what process or quality indicators might be quickly and easily collected which might suggest that progress towards the priorities was being made? And third, by what methods, with what frequency and by which individuals and groups, might the relevant information be collected?

The main priorities were identified as follows:

- the extent to which capacity building for programme implementation had started and the stage that has been reached;
- the extent to which steps had been taken to help the local community participate in the affairs of the local school;
- the efforts made to enhance girls' enrolment and that of disadvantaged groups;
- the provision catering for the diversity of children's learning needs; and
- the steps taken to provide effective external support for teachers to enable them to perform more effectively.

Indicators of Effective Project Management

The kinds of data required to assess project management included the following:

- nature of participation at different levels
- staffing of institutions
- physical infrastructure and any proposed improvements
- planning documents, site and supervision reports
- use made of non-personal resource inputs

These data were then used to construct the following indicators:

- the extent to which participation is seen as useful
- shortfalls of staffing compared to establishment, by programme functions
- levels of maintenance of non-personnel resources

- quality of plans

Indicators of Institutional Development

Indicators of linkages at each institutional level included:

- the areas of linkage (e.g. at the level of the State Centre for Educational Research and Training level, including training of Master Trainers, evaluations and research studies, production of materials); and

- the extent and effectiveness of the linkage process.

Three aspects of institutions were evaluated: training, materials development and ongoing professional support.

Training

Data were collected on the number of training courses and clients. This included subsets such as the training of Non-Formal Education/Alternative Education instructors, and in some areas, the training of Early Childhood Care and Education (ECCE) teachers; the timing of training (e.g. state training programmes intended for cascade might be initiated at times when training lower down the levels will occur at inappropriate times in the year); and processes of training.

The methods by which such data can be captured include survey and document analysis to pick up information on the number of modules offered, attendance numbers and timings of courses. The key indicators would be adequacy, organisation and relevance of the training process, for example, coherent sequencing of the training process (i.e. that the production of the required materials is in place before the training modules are implemented, etc.)

Materials Development

Basic contextual data include: the number of modules developed; the process of module development; and the availability and use of modules.

Methods for collecting this data include document analysis, use of focus groups, etc., discussions with recipients of training, participant observation, and structured and semi-structured interview techniques.

Ongoing Professional Support

Basic background data included: the frequency, number and purpose of visits to the school; then number and kind of follow-up activities; and the types of ongoing dialogue with teachers.

Key indicators were based on whether there is two-way communication or whether directives and training are delivered and no feedback is received; or whether needs and concerns of clients (generally teachers) were heard and responded to by the institution concerned.

A suitable timeframe for the simultaneous conduct of the evaluations involved collecting data, analyse and report back on the first component of the evaluation by the end of Year 2 of the project. This provided the appropriate stakeholders with some early information about the efficient implementation of training at a stage where corrective, formative and innovative interventions would be made. Capacity building and empowering of lower levels would be a guiding principle for data analysis. A further round of data collection and analysis will be made as a follow-up to subsequent interventions or ongoing implementation, sometime in the middle of the project. The exact decisions about when to initiate the follow-up study would be a state-wide decision. A final review would be made at the end of the seven year life of the project, with a particular emphasis, perhaps, on examining the sustainability of the institutions' effectiveness and quality.

Indicators of Community Participation

It was seen as important to find a way of collecting data on the part played by the community in the collection of data about enrolment and repetition, for example, through child-to-child monitoring, door-stepping, adopting of specific families/groups by the VEC - as an indicator of how much involvement there is at a real level from the VEC and other community members. The quality of community involvement in education, however, also requires indicators that register the satisfaction level of teachers and parents with the functioning of the VEC.

On an institutional level the setting up and functioning of VECs, the records of meetings, the incidence of discussions, the occurrence and resolution of critical incidents, have all been measured and taken as indicators of community involvement.

An indicator of growing co-operation and mutual support, for example, was the number of times that a teacher felt able to initiate questions or discussion with the VEC. Finally, a quality component was incorporated in the indicator itself: e.g. could members of the VEC discuss educational issues showing some understanding of the project aims in classroom processes?

Indicators of Access, Enrolment and Repetition

The collection of data through the school and pupil survey to supplement the initial

baseline survey and Project (PMIS) and EMIS data gave an aggregated source of information at the level of numbers of children from the village who are school-going children, the numbers of marginalised group children attending/not attending school and the number of drop-outs in a year. More detailed data needed to be collected from the population, and for this, there were some village-based exercises such as censuses and registers. However, households from which students come could also be the basis of a sample of households who send at least one child of school age (6-14) to school. Indeed, it is almost a PPS (probability proportional to size) random sample of households who send at least one child to school, but not quite because a household with, say 5 children may only be sending three of them to school. The correct weighting for the number of children in that age group whom they actually send to school, however, can easily be calculated from data collected at the interview stage.

It will be important to study a sub-sample of this kind for a number of reasons linked to the equity, enrolment and community participation objectives:

- the reasons for sending some or all children to school;
- the encouragement given to, or motivation of children to go to school; and
- involvement in school (PTA, homework, etc.)

The school and pupil survey will yield quantitative, factual data which will be able to be used in investigations of community participation. The triangulation of figures through interview, small studies, Participatory Community Appraisal (PCA) and a number of small case studies will add detail and reveal new and interesting questions, throw light on wider issues and corroborate data collected through the baseline survey and PMIS, EMIS. It will be possible to cross-reference across similar or different communities with similar project inputs, e.g. new school, trained teachers, and to note patterns of similarity or difference. This will provide the basis for further qualitative investigation.

Indicators of Teacher Training

From the point of view of the project, the quality of training can be thought of as evidenced in the changes that occur in teaching and teachers as a result of the training input. It is assumed that pedagogic and attitudinal changes in teachers can lead to improvements in children's learning, classroom interactions, and perhaps in terms of teacher attendance and retention also.

The in-service training of teachers can be divided into three sequential stages: preparation of materials, planning and implementing the programme, and the effectiveness of the programme in practice.

The Preparation of Materials Prior to Planning

Indicators of stakeholder participation in the preparation of materials proposed are:

- whether teachers have contributed towards the planning and preparation of materials;
- whether teachers have provided examples of good practice explicitly related to the training; and
- whether there have been trial runs of the training materials before planning, in order to smooth out any problems with the materials.

Organisation of the Training Programme

Assuming that basic data about the programme (for how many, how long, where it takes place, methods of delivery, content) has been collected, the kinds of indicators of effective functioning would be:

- whether actual training matches the plan and how the training programme is assessed;
- the extent of active involvement of participants including the existence of follow up sessions and their feedback; and
- the sustainability of the programme.

The Effectiveness of the Training Programme in Practice

The final outcome of training is its effect on classroom methodologies and children's learning. This will be considered in the sections on school processes and learning achievement. Specific training indicators relate to:

- whether teachers have changed their pedagogy or whether they mimic the model lesson when they are being observed and in all other lessons revert to didactic, front-of-class instruction;
- efficiency in terms of cost and the use of existing structures and trainers; and
- is there a significant transmission loss in the cascade model and does this vary between teachers and trainers?

2.3.5. The Overall Effectiveness of the Project

School-Classroom Processes

Key evaluation questions are raised in this area and these questions may,

themselves, inform the development of indicators of key classroom processes and understandings. Such questions include:

- What are the factors that make for differences between schools?
- What is perceived as a 'good school' by the community?
- How do teachers evaluate a 'good school' and what do teachers see as factors in school improvement?
- Do physical facilities have an important effect on school improvement?
- How are the effects of training shown in practice in the classroom, i.e. what does the teacher do that is different?

Answers to the first question would be desirable (world-wide!), and it is important to establish first what different stakeholders perceive to be a 'good' school. Only the last two issues are strictly relevant here, and the first has already been considered in another section (on the use of appropriate building/construction technologies).

Hypothesised indicators of a link between training and the classroom include:

- actions indicative of activity-based and participatory learning;
- use of resources and materials, especially those in local environment;
- lessons conducted in a gender sensitive way; and
- effective multigrade teaching - e.g. evidence of differentiated learning.

The Evaluation of Learning Achievement and Teaching-Learning Materials

(i) Non-Cognitive Learning Achievements

One of the biggest challenges under APDPEP is the broadening of the concept of what constitutes learning achievement. Children achieve many things in school; in a good school many of these achievements are in the psycho-motor or affective domain. Already Indian education places greater emphasis on citizenship, social awareness and parity within a secular state. How these aspects of children's learning can be integrated into learning improvement and progress measures, thus giving them validity and importance in the eyes of parents and the community will be an on-going concern throughout the project.

The 'ground work' for these studies is not yet in place. A sweep of primary school

practices in the non-cognitive domain is required. At the beginning of the textbook development component of APPEP, teachers were asked to send from their Teachers' Centres (TCs) examples of their work in Class 1. The response was overwhelming and substantially informed the work on the Class 1 textbook. A similar request for information about non-cognitive activities: craft work, work-related studies, e.g. agriculture, weaving, drawing, gardening, music, dancing, cultural studies and contingent activities such as scouts and bulbuls, might reveal a wide range of interesting activities.

Short studies using photographs, video and participatory discussion would suggest what is going on in these areas. It is unlikely that a *measure* of non-cognitive achievement could be devised, but teacher assessment or assessment by the community and reference to them in the teaching/learning materials would at least illustrated that there had been some impact of these activities.

(ii) The Development of Indicators for Learning Improvement and Progress

Indicators of progress at Class 2 and Class 5 are to measured by tests of literacy and numeracy. These measures will not be without difficulty. For example, a true measure of literacy must include primary children's ability in writing and oral language.

The development of assessment procedures for writing have been incorporated as part of training developments to include teachers learning about the processes of writing, and children being given the opportunity to develop writing so that it can be used to effectively express their own ideas and understandings. Practise in writing in the classroom, moderation of writing across schools at the TC and children learning to assess, edit and improve their own writing are an integral part of this process which will need to be supported by practical, experiential training for the teachers. Oral language, like writing, needs to be assessed on other criteria than 'correctness' and will be part of the same training package involving teachers and children in awareness of the importance of communication rather than correct copying and repetition.

Table 8: Project Specific Indicators for the Participation Objective in APDPEP

2. Making The Community Participate in the Affairs of the School	1. Involvement of the community in the planning process	· Evidence of education agenda as per record of meetings · Training of village core group in microplanning and school mapping process · Study involvement of women/ SC/ST in planning exercise · Analysing the constitution of the VEC in social terms. · Presence of women in VEC	1. Analysis of 5% statistical sample of schools through trained DIET Evaluators to seek evidence of participation. 2. Participant observations by NGO representative followed by report through the Block group to the DPO. 3. BRC/CRC Co-ordinators to evaluate social composition of core group involved in the planning process and also analyse the membership of the VEC.
	2. Support of the community to the school	· Evaluate financial support provided by community to the school in materials and labour · analyse donations received from community · time spent by parents in school during school activities · support extended to the school teacher	1. BRC/CRC staff can evaluate on the basis of VEC records 2. Process observation by resource persons involved in training VEC members. There should be a standardised reporting format. 3. Record perception of school teacher regarding assistance from community by interview method. It may be conducted by a DIET faculty/BRC/CRC staff.

	3. Participation of the community in the enrolment of children	Participation of community annually in the preparation of the Village Education Register along with the Teachers and the children of Class IV and V	1. Women's Self Help Group Leaders, Dwera volunteer to report on this activity to the Block Resources Group. 2. Women's Development Office in DPO to conduct some field studies in sampled villages. 3. This activity should be evaluated by CRC at their level and report on good and bad cases sent to BRC. BRC should share these findings with DPO. Some villages not doing well should be subject of an intensive study by external agency. Study to suggest remedial measure. 4. These studies to concentrate at the time of enrolment but there must also be a review every quarter at the CRC level. 5. Effective enrolment to be calculated by subtracting the number of children not attending school on 50% working days, from those shown as enrolled as per school register. School headmaster should make comparison with village education register to find out actual number of non enrolled children.

	4. Participation of the community in the activities of the school	· Number of functions held in school when parents were invited · Number of parents who attended these functions · Assistance provided in organising school sports, cultural meet, Bel Mela	1. School headmaster to submit report in format every six months to CRC Head. 2. CRC to visit school on function days and submit report to BRC. 3. Participant observation by resource persons in teachers' training for non cognitive areas. Their reports should be sent to DIETS. 4. DIET faculty to make comparative study of these reports.
	5. Formation of representative VEC	· Analyse process of VEC formation · Number of the core group members who took interest in the initial planning phase who have become members · Number of women, SC, ST represented.	1. BRC level resource persons involved in VEC training should be used as process observers. They should report in design format to DIET. DIET staff to undertake analysis of these reports and send small report every 3 months to the DPO in tabular form.
	6. Involvement of the school community in the school construction programme.	· Number of active members visiting school site during construction and recording their names in the site book. · Level of maintenance of accounts and its sharing in meetings.	1. Junior Engineer who visits site must check site book and report to Block Resource Group /DPO 2. School head master to report to CRC regarding school construction. 3. Retired teacher to make site visit and report to DPO. 4. Engineer belonging to another department to make site observation (he is external to activity)

2.3.6 Conclusion

This section has shown how, at a micro, evaluative level, it is possible to elaborate very large numbers of performance indicators. In principle, these could be modified by local communities to reflect their own concerns. But, there is clearly a danger that they will be taken far too seriously by the level immediately above the community, thereby distorting the local community's energy into meeting the performance indicators rather than achieving their own targets which are perhaps not easily expressed in such definitive terms. It might be better to experiment with the development of a pilot set of indicators covering only one aspect of community participation before extending to cover the wide range of issues potentially involved.

2.4 EDUCATION PERFORMANCE INDICATORS IN A NEW SOUTH AFRICA: A PROGRESS REPORT

This case study is intended to provide a snapshot of some of the developments taking place in the creation of education performance indicators in South Africa today. It is based on five days of interviews in Johannesburg, Pretoria and Durban in the week of the 7th April 1997.

If the only purpose of this case study had been to report on the **use** of education performance indicators in South Africa today, there would have been little to report, as the visit in April 1997 was too early. Thus, some of those interviewed reported that they were busy either creating the framework for data collection or in the data collection itself, in the absence of fully-specified indicators, let alone their use. The urgency of establishing a national data base of the country's schools has taken precedence over the development of indicators. And plans are being put into place for the development of the management capacity of appropriately *using* such indicators as are created in the longer term. In some ways this is perverse, as the realm of performance indicators is defined by the data available.

Many of the discussions held related to a theme emphasised in the first Kenyan case study, namely the **process** of deciding what information to collect. The political context in which such performance indicator schemes are developed was a focus of discussion, echoing a further theme raised in the first case study. What is significant is the conflation of purposes envisaged in performance indicator development in the discussions held, particularly as those with least political clout can easily have their information needs overshadowed. Thus, one of the focal points of this study will be to unpack the different uses envisaged for performance indicators at different levels of the education system and the data requirements for their construction, given the perspectives represented.

In what follows, first a matrix of different power constellations surrounding the different uses of performance indicators is portrayed as the different levels of actors are considered in turn. This rough matrix is not intended to be a definitive

representation of South African educational networks but it is intended to serve as a backdrop to some of the major issues concerning performance indicators that were raised in discussion. This forms the second part of this case study: how the roles of the different players are coloured by the influence of several educational initiatives already launched as well as legislation enacted. The third part of the case study discusses some of the indicators made possible by the current data collection exercises as well as those needing to be carried out to serve wider purposes than those already envisaged.

2.4.1. Framework: Levels and Purposes of Use of Performance Indicators

Table 9 very roughly sketches the main purposes and/or uses of performance indicators by different actors at different levels across the education system in South Africa today. The roughness of this portrayal should be emphasised, as well as the fact that it is a *current* picture; no doubt, many nuances at each level will be missed; and no doubt, over time, many of the boxes left blank will be ticked, as further responsibilities are undertaken at different levels. Nonetheless, the matrix does represent the slightly blurred snapshot taken in April 1997, albeit with somewhat artificial divisions of responsibility.[12] For instance, of course the national Department of Education is involved in planning. The provincialisation of responsibility, however, sets out a division of labour that accords the national Department the responsibility for monitoring and evaluating the provision of education, whereas the provincial Ministries are accorded the management and implementation of the programmes according to norms and standards developed at national level. Indeed, a recurrent theme raised during interviews was the "unfairness" of the provincial governments being held accountable for attaining standards, the prerequisites of which are not similarly mandated by the national government.

Thus, working from the top row down, according to the *National Education Policy Act*, 1996, the Minister of Education is responsible for the monitoring and evaluation of "the standards of education provision, delivery and performance...with the object of assessing progress in complying with the provisions of the Constitution and with national education policy." The directive principles of national education policy are set out in Section 4 of the Act and, among many other things, include:

- achieving equitable education opportunities;
- achieving the redress of past inequality in education provision;
- enhancing the quality of education;
- ensuring broad public participation in the development of education policy;
- achieving the cost-effective use of education resources; and

[12] In categorising the different purposes of performance indicators in Table 9, a division between *formative* evaluation and summative evaluation is made, the latter being subsumed under "accountability".

- achieving close co-operation between the national and provincial governments, including the development of management capacity. (*National Education Policy Act*, 1996:8)

As noted during discussions, this broad *national* responsibility for monitoring and evaluating the norms and standards of education was challenged in the courts. The national Department won, setting the momentum for carrying out this role. Following this first legislation, the *South African Schools Act, 1996* provides for "a uniform system for the organisation, governance and funding of schools". As will be detailed below, a new *Annual Survey for Schools* was carried out on the 22nd April 1997, thus providing the baseline for monitoring and evaluating national education policy. The norms and standards for funding schools in line with the *South African Schools Act*, 1996 are still under discussion by the Heads of Education Departments Committee (HEDCOM).

BC 99

Table 9: Matrix of Actors and Purposes of Performance Indicators in Education in South Africa

Actors/ Purposes	Monitoring	Formative Evaluation	Planning	Management	Accountability
National Department of Education	X				X
Provincial Ministries of Education	X	X	X	X	X
NGOs, Research Organisations		X			
Schools, Heads, Teachers and Governing Bodies	X	X		X	X
Parents, Students, Communities					X

Section 8 of the *National Education Policy Act, 1996* makes clear that the nine provincial Ministries of Education and Culture are accountable to the national Department of Education for the provision, delivery and performance of educational standards. It is their responsibility to remedy the situation if they do not comply with the Constitution or national policy. Thus, the provincial ministries must be engaged in monitoring, formative evaluation, planning and the management of education, falling short of the summative evaluation function left to the national department except with respect to their own constituencies.

It is worth mentioning at this point the 'sea-change' that was expressed as being necessary to begin to use performance indicators as tools of planning and management, to take two of the four functions designated at this level. With the exception of the matriculation pass rates in the past, indicators had not been published, and only direct input controls had been used. It is argued, there had been no 'need' for indicators to see what was happening in the system. The amount of education and training necessary to modernise educational management was not underestimated by anyone but, as with the development of performance indicators, where data collection has taken precedence, so is the placement and training of education officers required before a 'climate' of indicator use can be created.

In the next row, which is NGOs and Research Organisations, a variety of institutions are included, from education policy units, to university departments, to

non-governmental organisations involved either in school improvement activities or school effectiveness research. A lot of activity in this area is donor-funded, whether from within South Africa, for instance, through the National Business Initiative or the Independent Development Trust, or from outside, through DFID, CIDA, USAID or the many other donor agencies involved in educational aid programmes.

Formative evaluation is identified as comprising the main use of performance indicators for this level of activity. Of course, if one is talking about specific projects, all of the different purposes will be carried out. The focus of analysis here is, however, with the interplay not **within** the project, but between any projects and national or provincial educational development plans. Thus, the key linkage is what can be fed back, in particular, to improve the quality of educational programmes. Monitoring to ensure equitable distribution of resources is relatively simple. Evaluating the use to which different types and levels of resource are put brings us into the realm of school effectiveness and different notions of school quality, as it is understood and evaluated in different contexts.

The next category of 'actors' is at the level of the school. Putting heads, teachers, and governing bodies together in this category, however, does not necessarily mean that there is a uniformity of perspectives, despite the hope that all are co-operating in achieving the educational interests of the school. All but planning has been included in the use of performance indicators at the school level. This is because, at present, school-based management is not the reality that it might become. Thus, whilst everyday management is carried out, there is not widespread school development planning encompassing at the school level the sort of prioritisation of investments that are having to be done at the provincial or district levels. As per Section 20 of the *South African Schools Act, 1996* governing bodies are entrusted with the provision of quality education and are ultimately accountable both to the parents, students and the communities they serve, as well as being accountable to the provincial ministry and the national department. Thus, there is the need to use information (namely, performance indicators) for this purpose.

Finally, whilst parents, students and community members may sit on the governing bodies of schools and thus play wider roles in the school's development, their use of performance indicators outside such alternative roles, is judging the quality of the school. Thus, accountability is highlighted, whether from the perspective of the student's own evaluation of his/her performance or the parents' or the community's evaluation of the school's performance as a whole.

This matrix represents a range of very different perspectives and very different purposes of the use of performance indicators in education. Not all perspectives were canvassed in the limited time available in South Africa for this case study, nor were all purposes of performance indicators addressed. Nor is this matrix immutable. On the contrary, it is more than likely to change, particularly as data are collected, performance indicators created, and the management, planning and evaluation capacity developed across these levels in the many uses of

performance indicators. Against this backdrop, the next section describes some of the actors and their functions in the matrix as defined by several new initiatives in education in South Africa today.

2.4.2. Current Educational Initiatives

The national Department of Education has been preoccupied with the establishment of frameworks and mechanisms for policy change. Reference has already been made to two pieces of educational legislation which have set out the framework for national education policy and the governance and funding of schools. In addition, other significant initiatives have included the establishment of a National Qualifications Framework; a new, national curriculum, Curriculum 2005 (and related to these two initiatives, a South African Qualifications Authority); and the beginnings of a monitoring and evaluation system based on a new Annual Survey of Schools which will feed an educational management information system. Other initiatives which are still in progress, include the formation of an educational management development institute, a task set the interim committee established by the Minister, following the report of the Task Team on Education Management Development and subsequent legislation.

While indicator development as such has not featured prominently, the provision of a monitoring and evaluation system, and specifically an educational management information system (EMIS) has been a priority concern. The mandate for establishing an EMIS was given to the national Department of Education in the *National Education Policy Act, 1996*. A national EMIS steering committee was formed in June 1995 which presented a proposal to government in September 1996, following study tours of the management information systems in 20 countries and the deliberations of five task teams (Republic of South Africa, Department of Education, National EMIS Steering Committee (1996a; 1996b; 1996c)). The guidelines for the development of the new EMIS laid down the importance of consensus between the provincial and national departments of education, its facilitating role for these two levels of educational administration, that the system should be capable of providing data for policy making, planning, management and monitoring of the education system, that it should be demand driven, and that it should facilitate public access to education statistics and information (Republic of South Africa, Department of Education, National EMIS Steering Committee 1996b:2).

Parallel to the development of the EMIS, and in particular, the definition of a core data set for collection in the annual survey, another special survey was conducted of all schools in the country, the School Register of Needs Survey, designed to locate and register the physical resources and condition of all schools. The necessity of conducting this survey was the discovery of many 'ghost' schools and 'ghost' teachers, the inability physically to locate many schools and the uncertainty of their provisioning. Computerised maps of the data collected through a geographical information system (GIS) have been prepared, creating a provincial

level browsing programme enabling the viewer to choose which data s/he wishes to explore at the provincial or sub-provincial level.

Curriculum 2005 introduced in April 1997 builds on the National Qualifications Framework (NQF) established in October 1995 (*South African Qualifications Act, 1995*) whose objectives are to "create an integrated national framework for learning achievements", facilitating "access to, and mobility and progression within education, training and career paths." The new qualifications' framework has broken the previous, segmented pathways to exclusive, rather than inclusive qualifications.

Specific, context-specific learning outcomes, will serve as the assessed units, complemented by "critical, cross-field outcomes" representing much broader, educational goals. A continuous, formative assessment model will be applied to enable learners to progress at their own pace, and not necessarily through conventional, age-graded classes. Criterion-reference assessment, however, in line with the outcomes-based approach, will "underpin all classroom assessment, i.e measuring individual performance against the defined NQF standards." (Republic of South Africa, Department of Education 1997b:37). The South African Qualifications Authority (SAQA), appointed in May 1996 is responsible for formulating guidelines for accreditation, setting standards and monitoring achievement. External assessment will be done at the end of each level of the compulsory nine years of education, thus, at grade 3, grade 6, and grade 9. Only the final assessment will be used for selection purposes.

A Task Team on Education Management Development was appointed by the Minister of Education in February 1996. Its responsibility was to make proposals for improving education management capacity in South Africa, and more specifically for establishing a national institute for education management development. The Task Team produced an audit of education management needs and provision in each of the nine provinces, as well as its formal report published in December 1996, *Changing Management to Manage Change in Education*. Its recommendations build upon school-based management as a resourced and supported goal for bringing about quality, effectiveness and efficiency in the education sector, and are reflected in the guidelines for constituting school governing bodies laid out in the *South African Schools Act, 1996*. A task group has been set up to see to the establishment of a national institute for education management development.

Over this interim period in which frameworks, guidelines and responsible authorities have been defined for the educational reform, many non-governmental or research organisations have filled the gap in monitoring and evaluation, producing a variety of reports and studies examining schools, particularly as they are changing as a result of different interventions. Together with an examination of the development and potential use of indicators for the purposes identified in the matrix in Table 9, the third section of this case study looks at the role that this wider network of organisations can play and the design of their studies to fulfil some of these purposes.

2.4.3. Indicator Development and Use

Background Issues

It would be quite incorrect to say that education performance indicators had had no life in post-apartheid South Africa to date. Indeed, one of the most widely reported (to me) *uses* of a performance indicator was in the experience in the Western Cape of applying the pupil-teacher ratio (PTR) to equalise resource distribution across the provinces. The story tells as much as one needs to know about the importance of not using performance indicators as if they are technical, apolitical constructs. As the story was told, in 1994 the national Department of Education decided to equalise expenditure between the provinces and worked out budget cuts incrementally over a five year period, based on the pupil-teacher ratio. The Western Cape, having a very favourable PTR - about half as many pupils per teacher as the other provinces - managed to squeeze additional funds at cabinet level to get more for education and thus delay any teacher retrenchment in the first two years. In 1996, however, the province was forced to implement national policy, and retrenched several thousand teachers, only to be subjected to the political backlash of such an unlikely strategy for equalising teacher deployment. If only people moved as easily as the numbers in this exercise!

Four goals underlie the broadest use of performance indicators in South Africa today: equity, redress, access and quality. They are interrelated, as have to be the performance indicators that serve to monitor them. For instance, access is a starting point of the NQF, that there be multiple pathways toward multiple entry qualifications. If access, however, is only to poor quality institutions, then the significance of access is diminished. Ironically, in the case of sub-Saharan Africa, the lack of attention to quality has fed enrolment declines, reinforcing the view of the necessary expansion on both fronts.

South Africa is attempting a radical educational reform, compared to many of its neighbours. It is worth portraying the challenges involved in accomplishing these reforms.

When explaining the purpose of this case study during interviews, most respondents were quick to explain that there was no tradition of use of performance indicators. Indeed, what was absent was a culture of information use, full stop! There were two very different ideas underlying such statements. One view was that education had not been planned, in the sense of prioritising according to needs, so this would be 'news' to those on-the-job and a new area for those recruited into some of these new posts. Another was to explain how sophisticated South Africans had become during the apartheid era either of camouflaging themselves so that information reporting necessarily would be unreliable or that its danger had made them indisposed - not out of lack of experience, but rather out of mistrust - toward the use of information. There is therefore a challenge in generating appropriate data use at all levels of the system.

A related theme that emerged in discussions is also familiar in most countries: the divide between those with a background in information technology and those 'planners' and other administrators whose interests in the products, say, of EMIS, emerge from the questions asked of the system, however the data are generated; answers to these questions would not necessarily have been produced in the past with computers. Thus, there is often a divide between two, rather different traditions which requires some bridging to ensure that it is the planning questions, rather than the technology used that drives EMIS development and the creation of performance indicators. Nor is this a "one-off" situation, to be addressed only when discussions are taking place surrounding the selection of a core data set. The need to make and continually reinforce this bridge will remain as long as there are administrators trained without IT skills, and whose knowledge of the educational system is valued.

Finally, given the structures that have been created by the new legislation and the division of responsibilities, there needs to be a high level of co-operation and trust for the challenges of the educational reform to be realised. First, as noted above, the structures themselves create an imbalance in power and responsibilities: the national Department sets the norms and standards to be achieved, but cannot mandate the means of bringing these about. Second, the continual development of such co-operation and trust is dependent on relatively democratic feedback within the consultative networks that have been utilised to create this vision of reform, but which are now set with different responsibilities of bringing it into existence. No one interviewed denied that consultation has continued to play an important role in policy development in post-apartheid South Africa, but several pointed to the lack of accountability of representation from institutions to their own constituencies.

Indicators

School Register of Needs Survey

In addition to identifying the school, not only by name, and address but physical location, so that it can be positioned on a map, the following categories of information are collected, those relating to the students and teaching and non-teaching staff, the same as those collected in the Annual Survey for Schools.

General Information about school

- school type and level and type of funding and owner of premises
- number of pupils by class and sex
- number of teachers and types and whether subsidised or not
- number of non-teaching staff and types and whether subsidised or not
- platooning
- medium of instruction
- other uses of school facilities
- access roads
- sport
- improvements made to roads, drains, landscaping
- extent of resources provided by school by type

Physical Information about school

- administration areas specified by type and size
- specialist classroom areas specified by type and size
- instructional areas by size and whether prefabricated
- condition of buildings
- condition of school and equipment
- other facilities used for instruction
- telecommunication
- boarding facilities
- furniture, equipment and materials by type and provision
- water
- energy
- toilets and condition
- fencing

Two forms of output of summary statistics based on these data points were seen: (i) in a booklet produced for the first provincial workshops; and (ii) the provincial 'browser'. In the former most statistics were calculated by circuit, so one had bar charts of the number of schools, teachers, pupils, pupil-teacher ratio, pupils per classroom, instruction areas, pupils per toilet, staff per toilet, primary to secondary pupil ratio, shelters, instructional areas for special subjects, estimated shortfall of student, teachers' and administrators' furniture and equipment.

Further provincial level statistics were then calculated, illustrated in the form of pie and bar charts: schools by level, owner (by type) of premises, availability of telecommunication, water, water source by type, provision of fencing, power and energy supply, schools according to condition, funding, other facilities being used as well as other uses of school facilities, adequacy of provision of furniture, equipment and materials, and other resources.

It is clear that these statistics are essential for some of the initial, physical planning tasks at provincial level, particularly as much of this information was simply unknown. The indicators produced at the circuit level can help target resourcing by need where inadequate provision of basic teaching staff, facilities and materials is found. There is nothing detailed on the levels of qualifications of teachers, which is collected as part of the Annual Survey for Schools, but the Register will play an important role in providing a baseline for monitoring equity of resourcing and quality of physical inputs, outside more sophisticated notions of staff skills and training levels. The primary to secondary school pupil ratio will serve as an important, initial indicator of secondary school access, and as was explained to me for the Kwazulu Natal province, that was one of four indicators being used in the first instance, the others being learners per classroom; learners per educator; and educators per classroom.

Regarding the second 'product' of the School Register of Needs Survey, the 'browser', some of the maps produced to date for some of the provinces were examined.[13] These, like the summary statistical tables, were to be fed into the two provincial level workshops planned for each province. To date, what had been produced were maps of some of the summary statistics by circuit, e.g. the pupil-teacher ratio or the provision of particular resources.

What was interesting were some of the different reactions to the use of these data. For instance, one person at the district level for Gauteng, made the point that obtaining the raw data was much more useful than receiving the summary statistics that had been produced. This was because what was needed was to get their own people used to manipulating the data and creating indicators, so that they had both the familiarity with the numbers and an understanding of the meaning of the indicators. This chimed in with some of the verbal reports of the provincial workshops: that it was difficult to tease questions out of the participants, what they thought would generate the next steps forward in the computer mapping. What this suggested was a link to the point made above about different ways of 'knowing' an educational system and the bridge that needs to be made between those with an IT background, and those, for whom even planning is alien, no less the use of tables or maps of indicators. With the caveat that it may have been too early to judge, illustrative of this gap was the fact that no maps had been produced by the time of this investigation highlighting degrees of 'neediness' based on the variety of indicators made possible in the data gathering. This would seem an

[13] Not all of the data have been processed, so summary statistics, as well as the production of the 'browsers' for different provinces are at different stages.

obvious starting point for 'planning' questions. Maps are wonderful tools, particularly for overcoming the reluctance of many to engage in discussions concerning statistics. If those needing to use the maps are not able to query the data and extract information that is meaningful to them in their positions of responsibility, however, the attraction of using such high technology is diminished because of the disempowerment it can bring about. This underlines the importance of developing the capacities of those with either IT or education backgrounds so that they will be able to use such new systems to good effect.

National EMIS: Annual Survey for Schools

In the proposal for a national EMIS, it is noted that: "A system of indicators will need to be developed which annually tracks the progress of the education system." (Republic of South Africa, Department of Education, National EMIS Steering Committee 1996b: 13) Specific mention is made of the need to generate enrolment ratios and gender participation, both in terms of access and outcomes. Cohort flows to establish the internal efficiency of the system are also mentioned, and 'output' measures to ascertain quality improvement. Finally, it is suggested that "target groups will need to be identified to establish if the intended equity in delivery has been achieved." (Republic of South Africa, Department of Education, National EMIS Steering Committee 1996b: 13). As already noted, the mandate for monitoring and evaluation given to the national Department of Education in the *National Education Policy Act, 1996* is broad. This section will discuss the state of progress toward carrying out this mandate.

In the Annual Survey for Schools, more information is collected than is the norm for many other countries in their educational censuses. Specifically highlighted are the amount of income and expenditure data required from respondents as well as information on the composition and functions of the newly constituted governing bodies. Included in the categories of information on which data are to be collected are:

1. General Information

- Identification codes (including linkage with PERSAL (personnel information system) and examination centres
- Type of school (public/independent and former ownership)
- Grade coverage and classes/grade
- Multigrade
- Remedial
- Platooning
- Hostels
- Special School
- Learning Time
- Medium of Instruction
- Staff - state vs. other funding of educators, administrators and support staff
- Governing body since date, represented by..., functions
- Income and Expenditure (fees, fundraising, hiring, donations; salaries by governing body, texts, stationery, consumables, land, building, equip.)

2. Learners and Educators

- Learners by grade, race, sex, age; sex by grade; grade by home language
- Learners by secondary subject, by race, by sex, by grade
- Failures by grade and sex
- Repeaters by grade and sex
- New entrants by age and sex
- Transfers by grade, sex, and whether in/out of province

Teachers (linked to PERSAL) by sex, age, race, language, level of post and category, appt., experience, paid by, qualification category and type, subject specialisation, hours taught including remedial, subjects taught

This is an impressive list, and if the actual data collection can go as smoothly - and apparently reliably - as the pilot survey, then much of the mandate for monitoring and some for evaluation will be capable of being fulfilled. Linking the annual survey data with the recent population census clearly will make it possible to detail access. Similarly, the information gathered concerning learners at each school in the system will facilitate the construction of cohort flows. Although not without a battle, the racial categories were reintroduced in the new instrument, the arguments for facilitating the monitoring of redress and equity, clearly winning out to those concerned about not wanting to reinforce existing prejudices. To the racial categories have been added home language, so that it will be possible to track provision for different parts of the school population. Cross-tabulations of, say, achievement data, if these are truly linked to the schools' reports, and multilingualism, or particular home language dominance or percentage failing or repeating, will serve as a launchpad to further diagnosis of the system.

Some interviews emphasised that capacity building in the use of information at present is for redistribution, not for quality, so the development of indicators, in the first instance, will be relatively more straightforward (see below). If redistribution is meant to be on the basis of equity and redress, rather than a straight pupil-teacher ratio, as was used in the Western Cape, two alternative, extremely potent indicators would comprise some categorisation of educators' qualifications and/or educators' costs (roughly = salaries) by school; or government-funded vs. total pupil-teacher ratio, identifying those schools which hire many additional staff through the fees charged. Further, if reliable income and expenditure data by school are obtained, there can be further monitoring on equity grounds.

Again, to the extent that the culture of learning is restored and the education system becomes more stabilised, indicators of failures or repeaters by school and by age and gender and language group and/or medium of instruction similarly, will become potent indicators for closer examination.

The inclusion of information on the governing bodies is an interesting means of gauging the 'democratic transformation' of the education system at the lowest level. Given that the goal is for there to be school-based management, it seems a

very sensible area on which to collect data, in order to be in a position to target capacity building, as needed. The further informational requirements for parents and communities to become the ultimate judges of educational quality are, however, much greater than what is required to judge the abilities of the parents and communities themselves. Whether the governing body has developed a mission statement for the school or adopted a code of conduct for learners is something that can be verified, without too much obfuscation. Whether one can attribute to the school, rather than the pupils' own backgrounds, the quality of their academic achievement or other outcomes, however, is not as straightforward.

In the first instance, it is worth separating out different notions of quality. Quality can be and is often monitored on the basis of inputs, rather than processes or outputs. It should not be supposed that an annual educational census is capable, necessarily, of drawing out the causal links between these different factors. Yet, implicitly, our information systems are based on what are felt to be important constituent factors in monitoring quality (as well as equity and efficiency). If in the case at hand, the South African national Department of Education were only to concern itself with quality as inputs, redress would sound hollow. Comments on the development of some indicators proposed to be developed from the Annual Survey for Schools will be made in the context of recognising that it is important to monitor more than inputs bearing on 'quality'. Comments will also be coloured by our knowledge of a further constraint: the fact that our models of education are neither adequately addressed in annual census surveys, nor, one might add, in more sophisticated research designed to play out the causal links between inputs, processes and outputs.[14] In Annex 2A to this chapter further proposals for the development of indicators based on interlinked data sets are discussed.

Overlaying all the above attempts at indicator creation is the knowledge that outcomes (and thus, quality) will be expressed very differently from the relatively simple matriculation passes, as different national assessments begin to be made at the two new lower levels, grade 3 and grade 6, and the use of continuous assessment at the school level, requires a very different understanding of 'outcomes' from what has been used to date.

The norms and standards for school finance are in the process of being formulated. The national Department is caught, however, much like a donor - in not being able to specify how much a provincial ministry should spend; instead, it is likely to be left with 'conditionalities', e.g. proportions of the budget to be spent on *x* and *y*, respectively. These conditionalities, like any indicators used for accountability, will be contested by the provincial ministries if the basis of their determination is in any way questionable.

---

[14] Educational researchers would be ecstatic if they were able to account for as much as half of the total variance in educational achievement in a population, after controlling for family influences, leaving much room for factors clearly out of the reach of either the national Department or the provincial ministries. Yet, the latter are meant to be accountable for such quality outcomes.

Specific project evaluations feed the formative evaluation role of implementing institutions. They also raise questions about macro vs. micro evaluations of *quality*. Discussions with the English Language Educational Trust (ELET) illuminated one aspect of the changing design of such evaluations. It was noted, for instance, that in the evaluation carried out of DFID's Primary English Teaching for Rural Areas project (PETRA), that inputs (or you could say, processes) are the observations, not the outputs in the sense of the students and their learning. One is observing teacher behaviour, so one is looking at the quality of learning only indirectly. Examples of the observations recorded will be familiar to many: recording the use of materials by learners and by teachers; the grouping of learners; the questions asked by learners and by teachers, etc.

Similarly, in a collaborative evaluation effort with the USAID Improving Educational Quality (IEQ) Project team, the main questions asked concerning the impact of ELET training at the classroom level were:

1. In what ways do teachers with different levels of training teach differently?

2. In what ways do learners in classrooms taught by teachers with different levels of training participate differently?

3. In what ways do the classroom learning environments of teachers with different levels of training differ?

4. What is the relationship of other variables such as education, teaching experience, age, and gender on teaching, learning, and the classroom learning environments?

Learner participation, interactions with the teacher and other learners, and use of materials, for instance, define the learner outcomes identified for observation. What results is a rich data source on the changed classroom experiences from both the teacher's and learner's perspectives, but missing is the more direct evaluation of learning outcomes.

Such monitoring and evaluation of change at the level of the school and/or classroom, while serving the formative evaluation needs of those designing in-service training programmes and materials, falls short of the needs of those focused on whether learning outcomes are thereby improved as a result of the enhanced classroom experiences brought about. This is quite characteristic of school improvement projects, falling short on the evaluation of student outcomes.[15] The draft logframe for a new DFID education project in the Eastern Cape (ECSIP), recognises this. Built on a baseline study of pupil-level literacy and numeracy, evaluation is made possible of enhanced student outcomes.

[15] See Dalin (1994), especially p.181.

111

2.4.4 Concluding Remarks

There are many other examples of evaluations on which one could draw to illustrate the different data and performance indicator needs at different levels and for different purposes, as well as what, in effect, are very different starting points, whether from a systemic, macro-level school effectiveness orientation, or from a more micro-level, or at least school-based, school improvement orientation. The provincial Ministers of Education are concerned with how best to allocate limited resources, given the challenge of meeting the inclusive goals of the new education system. The project manager's concern, or the school head's concern, is whether the project inputs, e.g. in service training, have resulted in changed classroom behaviour. Ultimately, they are concerned with school quality, particularly, if they are to be made accountable increasingly to performance indicators designed to evaluate the quality of education in terms of learner outcomes. For the time being, however, monitoring and evaluating the effectiveness of the processes or inputs has not had to run the full gamut to include learner outcomes.

In the conceptual framework portrayed in Chapter One the problems of achieving consensus in choosing indicators for different audiences and different purposes was raised. This snapshot of South Africa in early 1997 illustrates the challenge of addressing the diversity of purposes and perspectives in such indicator construction. It is worth pointing to a fundamental, philosophical divide that underlies some of these different perspectives. It is described clearly by Carrim and Shalem (1996:2-3):

> In general, the delimited research emphasis on uniformity and homogeneity across schools leads to generalised, macro formulations and technicist solutions that assume schools are rationally organised systems and that their experienced problems may be remedied predominantly by monetarist interventions.

They go on to write:

> While the quantitative-based model of efficiency can measure, compare and calculate test scores against the background of classroom and teachers ratios and other inputs; can prescribe organisational procedures for an efficient management, what it can not do, however, is to tap into the interests and perceptions of the students, parents and teachers - the main social actors of the educational process. (p.18)

The centre/s of policy making must appreciate and have knowledge of classroom realities and the classroom realities need national (and global) grounding. The danger that South Africa faces is not really as different from other countries as some would have us believe. Should the urgency of policy making take such precedence that the informational (no less resource) requirements of policy implementation become eclipsed, then the distance that has been created between the national Department of Education and the provincial ministries, rather

than being bridged through trust and co-operation, increasingly will become arenas of power contestation. Thus, provincialisation runs the risk of undermining rather than enhancing stability. The challenge of identifying and appropriately using performance indicators in the new South Africa demands sensitivity to the different perceptions of stakeholders and policy makers at the different levels portrayed in Table 9, linking their distinct purposes of monitoring, formative evaluation, planning, management and accountability.

ANNEX 2A

Further Proposals for the Development of Educational Performance Indicators in South Africa

In a paper commenting on the national EMIS proposal, Dr. Luis Crouch, a consultant working within the national Department of Education, has written about the need to link the examinations data base with the annual survey data:

> It will be vital to being able to defuse all kinds of fuzzy debates about "quality" and it will also be vital to being able to identify schools or districts engaging in highly productive practices, and being able to generalise and reward those practices. (Crouch 1996: 10)

He develops this idea further suggesting the interlinkage of socio-economic data from the census with the distribution of resources (from the annual survey) and examination results, indicating the possibilities of identifying schools that are "outperforming others relative to a) funding and b) socio-economic characteristics." (Crouch 1996:14)

There are a number of strands in Crouch's argument which require further examination in the context of the potential development of indicators. The degree of 'weeding' of non-exemplary schools that can be performed, or as one might like to conceive of it, alternatively, as 'cherry-picking' of exemplary schools, on the basis of the interlinked data, will be determined by the variation in recruitment between schools within any one census district. If there is a lot of variation *within* the census districts, then, using examination results, one is more likely to pick out those schools creaming the best students from the district than those which are exemplary, because of what happens at school. Although the data were not available with which to judge the extent of variation, given the vision of school-based management, local accountability, and ultimately, 'choice' which lies behind the agenda on governance, even if, at present, there is very little variability, this is most unlikely to remain the case. Thus, attempting to do what is, ultimately, a very rough and ready cherry picking, is likely to confound performance due mainly to student background than to school activities. Thus, the 'fuzziness' Crouch wants to avoid by such interlinkages, is likely to crop up in another form, namely these poorly specified indicators.[16]

The second point about developing such indicators relates to the different purposes for which they are more than likely to be used, particularly in the absence of other 'quality' indicators. Although Crouch is advancing the use of such indicators for planning and management, and thus feeding a formative evaluation

[16] Indeed, this is precisely what was intended to be done in Namibia, utilising such interlinked data sets. The point being made, however. about adjacent schools exhibiting very different intakes within a single census district, was picked up early on when it became clear that such analyses were confounded by such variation. (See Namibia Ministry of Education and Culture, et al. 1994 and critique by Riddell, 1997).

agenda, there is little that would stop them being used for quite another agenda, accountability stakes, particularly as they look so 'reasonable'. One has only to look at the experience of the United Kingdom in this respect to understand how any sort of league tables, notwithstanding the caveats made concerning their use, become adopted and used in public debate.

Crouch also puts forward some suggestions for efficiency indicators, including learning output/input usage. This particular indicator is also predicated on the interlinkage of examination results with the annual school survey data. Although the argument used here, that "we will certainly never achieve efficiency if we only measure inputs and superficial indicators of quality such as pass and repeater rates" (Crouch 1996:11) is correct, these hardly account for "real measures of efficiency", given that the "learning output", yet again, will be confounded with the students' home backgrounds. Clearly, one wants to be able to judge whether there is movement toward the goals of redress and social justice, so to some extent, merely producing indicators that can highlight disparities in resource distribution will be of some value. What is tendentious, however, is trying to judge the relationship between differential resourcing and cognitive outcomes. If one found that two schools produced the same 'output' for very different levels of resourcing, the questions that one would ask would relate to the differences between the two schools that account for such different 'productivities'. Some of these differences would relate to the students themselves; others would relate to the measured 'inputs', some 'resourced', but others not, and the great bulk of differences would be unmeasured and unmeasurable, judging from researchers' successes at explaining even a simple majority of the variation in outcomes! This litany is precisely what many school effectiveness researchers have been asking for several decades, refining models to portray more vividly the reality of the school and the classroom. Therefore, it is with some surprise that in the context of thirty years of education production function calculations, Crouch's paper appears to be suggesting that an 'answer' has been found in a simple efficiency indicator. Even were such an indicator to be utilised alongside others, it would be seriously flawed as a complementary tool for policy analysis.

What, however, is exciting about the Crouch paper is that it is grappling specifically, with the interface between macro and micro-level quality indicators, and straddling the formative evaluation/accountability stakes divide by suggesting indicators, on the one hand, that would feed back information for improving the system, while at the same time, be capable of being used as a means of monitoring educational change. Relatedly, the paper is very much alive to the problem of regional aggregation of statistics and the clouding of variation within as well as between regions (Crouch 1996:17), even if it seems not to have addressed sufficiently the aggregation problems of applying census level data to schools' outcomes, particularly when it will all have to come down to the level of the student in the end, if one is to see changed outcomes.

CHAPTER THREE

BEYOND EDUCATIONAL PERFORMANCE INDICATORS: MOVING TOWARDS OVERALL SOCIAL INDICATORS

CHAPTER THREE

BEYOND EDUCATIONAL PERFORMANCE INDICATORS: MOVING TOWARDS OVERALL SOCIAL INDICATORS

3.1 INTRODUCTION

This chapter moves beyond the discussion of educational performance indicators, to consider both overall system of social indicators and ways of implementing them, both at a macro and micro level. Methods of assessing impacts of education upon other systems are discussed, in particular, the impact of education on health and employment.

In order to understand the rationale for socio-economic measurement, it is important to review it's history. In the first two sections, the origins and purposes of socio-economic measurement in the sixties and seventies are reviewed. It is concluded the major problem then was with the quality of the data available.

There have been several attempts to construct composite quality of life indices ranging from the PQLI in the 1970s to the UNDP's Human Development Index in the 1990s. These are analysed and rejected in section three. On this basis, the next three sections analyse the principles and criteria for constructing a flexible framework of social indicators and exemplify this in terms of assessing the impact of education on health and employment. Recognising the difficulty of implementation, the requirements and principles of monitoring basic needs at the local level are detailed in the final section.

This chapter is supported by several annexes which detail further background issues: opportunity cost and valuations; technical problems in developing indicators and concepts of employment and unemployment; and experience with local planning.

3.2 BACKGROUND

3.2.1 Origins and purposes of systems of socio-economic data

Although the recording of socio-economic data systematically can be traced back at least to William the Conqueror's Domesday Book, modern interest in reporting on (the lack of) social progress can probably be attributed to people like Quetelet (a Belgian statistician working in the middle of the nineteenth century) concerned to document the downside of industrialisation. Moreover, whilst the art of social measurement developed either as a consequence of increasing trade (Petty 1690) or as offshoots of the development of administrative systems required by the embryonic welfare states (Quetelet), or concerns of liberal reformers to measure poverty (Booth, 1890), the attempt to develop socio-economic data **systems** is

recent: it is a delayed consequence of the development of the national accounts during the 1930s and 1940s.

During the twenty years that followed the Second World War, the success of Western capitalist economies in raising living standards meant that, although there was some interest in modifying national accounts in order to provide a better picture of economic welfare - for example, Net National Welfare (NNW) - there was little specific concern over social data. Indeed, the only systematic attempt was that promoted by the United Nations, the System of Social and Demographic Statistics (SSDS). It was proposed to monitor the progression of people through their life cycle in a systematic fashion. Not only did the system imply very complex and exhaustive data collection, however, the reliance on the length of time spent in a status as the single dimension for assessing the value or weight to be attached to that status was very limiting.

3.2.2 The Social Indicators Movement of the 1960s

Social Indicators to Monitor 'Progress'

The resurgence of interest in a system of socio-economic measurement across a broad range of subjects in the 1960s can be traced to a number of factors coming together:

- The re-documentation of poverty in several Western European countries and the parallel questioning of trickle-down theories;
- A growing concern about the environmental and other social costs of economic growth in general and more specifically, the upheavals generated by the space programme in the United States;
- The emergence of a youth culture questioning the purpose of economic growth.

Hence, there emerged a concern to develop methods to monitor the quality of life in order to provide a measure of social progress. In particular, emphasis was placed on measuring the *outcomes* of social programmes rather than the financial *inputs* or the (usually administrative) *activities* associated with any specific programme. The problem then becomes one of specifying exactly what are the outcomes that should be measured.

Social Indicators for Managing Change

The concern with managing social change is one which has been clearly expressed by proponents and critics of the 'social indicator movement' (so called by many authors in the 1960s) alike. Bauer, working for the American National Academy of Science on research funded by NASA, was concerned with anticipating

the 'second order' effects of innovations such as the space programme, which may be more important than 'first order' effects. By this, he meant not only sensitivity to feedback but prediction of the consequences of feedback, which he saw as one of the main purposes of social indicators (Bauer, 1966).

Wiles (1971) was more specific about the kind of social change involved, seeing 'crisis prediction' as the goal behind the demand for new kinds of social information. This objective he condemned on two grounds. Firstly, because of the 'totalitarian implications' of 'efficient social information' which will 'stabilise whatever group is in power' by enabling it to be more flexible, buy off opposition more efficiently, increase its ideological hold, etc. Secondly (and 'fortunately'), crisis prediction never is efficient, whether one considers specific crises or looks at major underlying causes (particularly technological development). It is clear therefore that, for both these authors, the purpose of social indicators is maintaining social control, seen as threatened by a rapid rate of social change, complex and so far unpredictable, brought about by technological innovation.

Hence the interest of social indicators from the point of view of social control by the state is clear; if **all** the consequences of technologically innovatory programmes can be predicted, then programmes can be chosen not only according to their estimated profitability, but also according to their value in terms of preserving the system. Of course the implementation of such a system raises a number of problems. There are particular difficulties with measuring social change compared with measuring a stable economic system. Thus, not only is there a lack of any agreed theory of social change while the economic system comes - as it were - apparently readily quantified, it is often argued against social indicators that they are an attempt to quantify the unquantifiable; and the lack of a common unit of measurement makes it difficult to weight the social indicators so as to produce one, or a small number of indices (see Sheldon and Moore, 1968.)

An Uneasy Balance

The distinction between whether measures are to be used in the research and evaluation context or in a planning mode is important: for the planning purpose, almost any estimate is better than no estimate at all; while for research and evaluation purposes, the premature publication of proxy data may be counter-productive (Murray 1991).

But, whether the aim is reporting on the outcomes of social programmes, quality of life, social progress or simply social trends, or in enlarging social control *à la Bauer*, the choice of what components of well-being are to be measured by indicators is essentially a political one. Obviously, any combination of these indicators via weights that would themselves be the subject of debate - would be at least as politicised.

3.2.3 Approaches: A Thousand Flowers Bloom

Since the 1960s, there has been a flowering of different methods of measuring the quality of life in industrialised societies. A classic division might be in terms of methods, theory and policy. Among the methodologists, there are those who have followed the survey route: there have been large number of surveys to document the 'objective' living conditions of individuals and households ('objective' in the sense that the measurement is made by someone else). Equally, there have been a large number of attitudinal and/or opinion surveys and, sometimes, more systematic attempts to assess 'happiness' or 'satisfaction' in terms of scales. Other methodologists have directly addressed the problem of measuring the quality of life through a variety of valuation techniques. The theoreticians, unsurprisingly, have been less concerned with data collection procedures and more concerned with monitoring collective attributes of a society - sometimes seen as national characteristics - such as the level of autonomy, political participation, etc.; or with elaborating a model of social systems in order to identify its key attributes. The 'policy' group is concerned with providing simple policy tools - and so have focused on possible ways of modifying GNP to provide a better measure of (economic) welfare; or on combining 'objective' indicators (usually) into a composite in parallel to national income per capita.

There have, therefore, been several strands in this history often with very little communication between them. For the purposes of this review, the division into methods, theory and policy is not very useful because we are searching for social indicators that are theoretically grounded, for appropriate methods of data collection, and for relevant policy applications. Instead, we discuss the various strands in two main groups.

There are those who emphasise the importance of a uniform method of valuing welfare. Much of this work has been based on the financial nexus and have led to proposals either for methods of extending GNP to reflect better economic welfare; or for ways of valuing other, currently non-monetarised, components, using the measuring rod of money. There has also been interesting work, however, carried out using time as the basis for valuation and these are also discussed in detail in the Technical Appendix 1 under the general rubric of 'opportunity cost'.

Others have argued that we must escape from a system of data which are independent of the national accounts or of opportunity cost; and those working within that paradigm form the 'social indicator movement'.

In both cases, there has been further tension between the emphasis on the techniques of data collection and analysis, and on the validity of the data being collected for measuring the phenomena concerned.

3.2.4 Factors contributing to the decline of social indicators

Several factors have been identified as contributing to the decline in social indicators:

- There was a 'tendency for indicators to become vindicators' (Bulmer, 1990:410) and for the reports to be 'rather bland compromises, deliberately presented without text that might link the data to policy' (Innes, 1990:430).

- The system became divorced from the policy context, focusing on the measurement task, often to the exclusion of the political and institutional one (Innes, 1990:431).

- There was the lack of general social scientific theories to measure the theoretical constructs.

Most commentators - including the INES programme - have seized upon the last explanation. Carley (1981) argued that social indicators require too much time for elaboration, this being a reason for their decline; but it is the organisation of data systems which takes resources and time, not the definition of social indicators.

Carley's ahistorical argument ignores the fact that social indicators were developed in reaction to the perceived chaos of the 1960s, environmental concerns - both nuclear and pollution - and incomprehensible youth. It was because these threats had receded towards the end of the 1970s (they didn't disappear), or had been replaced by other concerns (e.g. over energy) that people lost interest.

The basic problem of understanding the social benefits to be derived either from corporate or government activity remains, however. While temporarily replaceable by crude indices based on recorded financial transactions such as GDP, serious analysts have always searched for some other system of measurement.

3.3 SOCIAL INDICATORS SYSTEMS

3.3.1 Early Examples in the Seventies

Three early examples illustrate different approaches to the wider policy integration intended by the development of social indicators.

Net National Welfare Measurement Committee of Japan

Established in May 1971 by the Economic Council of Japan, the NNW Measurement Committee proposed an index of "welfare national income" or "net national welfare" which would "allow for ... various plus and minus factors that do not appear in present GNP figures (OECD, 1976:35). They included the following

items in NNW: government consumption; private monetary consumption; services from government capital stocks; services from personal consumer durables; leisure hours (at average wage rate) and non-market activities such as housework. Subtracted were environmental maintenance costs and pollution reparation expenses as well as losses due to urbanisation in commuting and accidents; and net investment.

The initial calculations by the NNW Measurement Committee suggested that the average annual rate of growth in NNW was some 20% below that of Net Domestic Product (Economic Council of Japan, 1973). The NNW - probably for that reason - was never adopted as a measure of welfare.

The Physical Quality of Life Index (PQLI)

The Overseas Development Council looked for a measure to assess nutritional success in very poor countries. They said "The ultimate test, obviously, was the physical quality of life achieved." (Morris, 1976). A useful measure should meet the following criteria:

- the measure should not depend too heavily (if at all) on market performances (or on GNP data);
- the index should avoid measures that assume that LDCs will inevitably develop along lines followed by developing countries;
- the index should avoid measures that are excessively ethnocentric;
- measures of performances probably should not be based on absolute minima;
- the index or indicators must be sensitive to distribution results; and
- the measure must be simple.

Morris proposed the PQLI, defined as the average weighted rank of each country in respect of life expectancy, infant mortality and literacy. The measure satisfies the first and fourth criteria and, with a sufficiently broad definition of literacy, might well pass the second and third test. But it is clearly not very sensitive to the extremes and its operational characteristics - the way it reflects any specific situation - are not simple to understand, both because of its relativity and because it combines incommensurable items.

This latter point - the way in which one overall index tends to obscure rather than clarify issues is, of course, one of the original objections to the GNP measure being used as an index of (economic) progress. Proposing the PQLI - and it's latter-day derivative, the Human Development Index - is rather like jumping out of the frying pan into the fire. (see Section 3.4 on composite indices).

Active Life Sequence.

Seers (1975) has proposed that the well-known indicator of life expectancy should be extended to calculate expectations of the time spent in various 'life states' - for example, attending school, working, retirement, and so on. Later he argued that this technique would facilitate a statistical description of a community in terms of the typical life cycle of members of that community (Seers, 1979) based on the system of social transition matrices developed by Stone (1972).

Notwithstanding the basic practical problem of collecting a detailed record of movements by members of a population from any one state to another, the theoretical presumption of this approach is that the most important aspect of each and every life state is the length of time spent on it. Although the basis of some indices[17], it actually seems rather implausible as a *general* rule. It is indeed important to recognise that time - like resources - is limited and so has to be allocated and shared. But our concern with policy integration is wider than this.

Of the three approaches, PQLI is obviously far too simplistic to be useful for tackling the kind of problems that arise. Moreover, it has no theoretical foundation and it is therefore surprising to see that such a "stitched-together" measure is proposed in the academic literature as a 'useful' index (Cereseto and Weintztein, 1986) - and of course, as mentioned above, has spawned the HDI (see also 3.4.1). In contrast, both the NNW approach and the Active Life Sequence technique are interesting partly because they have some conceptual foundations but they are, in consequence, limited in their applicability. Thus, the NNW approach depends on being able to assign a monetary value to each of the dimensions included, with attendant problems of valuation (see Annex 3A); and the Active Life Sequence approach considers only the allocation of time to various activities and not the value of that time (see also Annex 3A). Nevertheless both are theoretically based and may shed light on some of the issues involved in integrating policies and programmes.

3.3.2 Basic Needs Approach

The notion of a hierarchy of needs - and therefore of basic needs - was developed by Maslow (1944). An emaciated version of this idea was taken up in the basic needs approach to development as defined in the Programme of Action at the 1976 ILO World Employment Conference. The 'basic needs approach' aimed to take account both of what goods and services are available and who were the beneficiaries in terms of consumption. Thus the 'definition of a set of basic needs, together constituting a minimum standard of living, would at one and the same time assist in the identification of these (poorest) groups and provide concrete (production) targets against which to measure progress' (ILO, 1976:31). Basic needs were taken to include two elements:

[17] For example, some epidemiologists advocate calculating life expectancy without disability; and there is always the QALY story (Weinstein and Stason 1997, Carr-Hill, 1989, 1991).

( ) certain minimum requirements of a family for private consumption, as well as certain household equipment and furniture; and

( ) essential services provided by, and for the community at large, such as safe drinking water, sanitation, public transport, health, educational and cultural facilities.

They argued that the following basic needs should be satisfied for everyone:

- security, food and water, clothing and shelter, sanitation (the survival needs);
- access, knowledge, mobility and skills (to function in society);
- quality, justice and self-reliance (to express a fundamental identity).

In contrast to previous emphases upon growth maximisation and industrialisation, the objectives were defined in physical terms. Overlaid on these requirements was an emphasis upon participation:[18]

> 'The main thrust of a basic-needs strategy must be to ensure that there is effective mass participation of the rural population in the political process in order to safeguard their interests.' (Sheehan and Hopkins, 1979:59)

Critique and Counter-Critique

The Basic Needs Approach has fallen into disrepute; critics variously argue that it is utopian, romantic, populist, anarchic, inefficient, slow, and that it obscures the fundamental international problems that frequently underlie poverty. Instead, they point to the need for freer trade or for fair prices (take your pick) in order to resurrect growth-oriented development (see Wisner, 1988).

But the special feature of the basic needs approach was the very belated recognition of the fact that welfare must be measured in concrete terms for each individual prior to homogenisation in monetary or any other units and prior to any aggregation across people in terms of a 'growth' index. No monetary measure of income per capita can ensure that essential goods and services are produced in the right quantities at the right time and actually reach the right people. Just as the policy implications of the approach were soon perverted (see Wisner, 1988), however, so were the statistical implications.

[18] This emphasis on participation has, of course, recurred in the EU concern with the exclusion of minorities from effective participation in society.

Despite the initial affirmation of the importance of identifying basic needs in physical quantities, those applying the approach in the latter half of the 1970s were eager to find a one dimensional yardstick (Hicks and Streeton, 1979) and sometimes a monetary equivalent. Indeed, the empirical papers in *Employment, Growth and Basic Needs* (ILO, 1976) are expressly designed to show what rates of growth in the measured economy would be required to satisfy a given collection of basic needs, leaving aside all kinds of questions about changes in the structures of the economies.

The Problem of Precedence

Maslow, based on then current theories of self actualisation, had promulgated a strict priority ordering among the different needs. In contrast, liberal eclecticism led the ILO to make no judgements at all about the relative importance of different needs (although of course, they crept in via the calculation of a Basic Needs Income). Yet, from any given perspective, there is an order of priority among 'needs': the Aristotelian and Marxist will tend to emphasise indicators of activity, whilst the Platonic liberal will tend to emphasise indicators of status, and so on.

Similarly, the development of a set of performance indicators is context dependent and more generally, the assessment of the impact of a set of policies is always from a particular perspective. This situation has led to proposals for a theory of social reporting (Johansson, 1976) so that one could purposefully choose indicators which are relevant for human development (Miles, 1985), taking account of the inter-linkage between areas. In this case, we are examining the potential for indicators of performance in basic education in developing countries and the impact of that education upon other sectors.[19]

3.3.3 Moving Towards the Individual

This emphasis on physical quantities rather than aggregate indices, however, led to measuring an individual's met needs, and thus, an individual's subjective needs were added to the otherwise, physical list of basic needs.

Subjective Happiness/Satisfaction

One such set of measures is concerned with deriving satisfaction measures. A systematic approach to measuring happiness and/or satisfaction has been developed by the Michigan school, the major exponents being Andrews and Withey (1976). They argue, on the basis of small-scale survey work, that several domains contribute to the final outcome of happiness, and that responses to questionnaires about satisfaction in respect of each of these domains can be used to generate a

[19] We might have been, for example, more concerned with assessing the impact of childhood health on learning or the impact of the labour market on the demand for education.

125

happiness scale. (The current version of this approach is the series of EuroBarometer surveys.)

The problem is that, if direct questions about satisfaction are asked, nearly everybody responds satisfied, and a large proportion 'very satisfied': this is partly because responses appear to measure social norms (of the 'can't complain' variety) rather than self-ratings of well-being.

Citizen Surveys

Another, Scandinavian approach (Johannson, 1976) has been to argue that objective data on living conditions should be available to all citizens as a pre-requisite in a representative democracy. While clearly objective data have to be included, reliance on citizen reporting raises all the usual technical problems of social surveys.

Indeed, the crucial issue is whether several of the important outcome indicators can easily be collected via the survey method of either type. This is not only because of the difficulties over response rates, etc., but also because the measurement of some of the most crucial characteristics of modern society is not amenable to the survey method.

For example, during the 1980s in the UK, an indicator which was used as a highly political football, was the numbers of deaths from hypothermia during the winter months; yet dead people tell no tales, even to a survey methodologist. Other difficult indicators that have been attempted have been the number of individuals in precarious (or without) housing; a further type of data which may be seen as an important monitor of the extent to which basic needs are satisfied is net nutritional intake. Again, data on activities or counts of events over a period raise problems of recall that can only potentially be solved via an expensive longitudinal panel design.

3.3.4. Conclusion

There are a variety of approaches and theories of what constitutes well-being, and while it might be difficult to claim that one perspective is superior to another, it has to be recognised that different perspectives generate different sets of priorities.

At the same time, what appear to be radically different viewpoints, tend to converge on a similar list of main constituents, while of course varying in the way these are organised, the emphasis, weight or rank given to each, and so on.

Table 10: Possible Framework of Social Concerns

Proposed Social Concerns	Content (for illustration)
HEALTH	Length & Health-Related Quality of Life Children's Future Development
LEARNING	Experience of School Levels of Ignorance
HUMAN ACTIVITIES	Use of Time Quality of Activities
WORKING CONDITIONS	Security Quality of Working Life
NECESSITIES	Basic Needs Fulfilment Poverty Lines
PHYSICAL ENVIRONMENT	Overconsumption of Energy Pollution
	The Family Victimisation The Wider Community
RELATING/ SOCIAL ENVIRONMENT	The Family Victimisation The Wider Community
PEOPLE & THE LAW	Restrictions on Movement Interference with Liberty

The OECD programme was intended to provide a 'system' of indicators to be used in the reporting of social issues and the formulation of social policy. In practice, however, the goal areas they proposed such as 'health', 'individual development through learning', 'employment and quality of working life' etc. have been treated as being in almost one-to-one correspondence with the 'appropriate' ministries of health, education, employment etc. Moreover, the enumeration of the indicators themselves does not force the necessary collaboration between different administrative units.

While precise measurement of these various aspects would require new data systems, a major focus should be on improving the comparability and coverage of data collection systems that are already in place. Within such a broad framework, however, it may well be sensible, in any one period, to choose a few key indicators that could be the focus of efforts to improve data collection systems. Table 10 details some of the common social concerns that could provide a possible framework for the creation of social indicators. The next section reviews some of the more recent composite indicators, while in the following two sections, the problems of defining indicators of impact on education on other sectors are considered.

3.4 COMPOSITE QUALITY OF LIFE INDICES

The purpose of composite quality of life indices sometimes is to provide a better measure of national or regional performance with regard to human welfare and development - what one might call real poverty reduction. They are usually seen, however, as a complement to, rather than as a substitute for GNP, and are constructed from among components of a broader system.

In fact, there have been a whole series of attempts to construct such composites both in developing countries and in industrialised societies. The earlier PQLI has been discussed above. The review here will only consider a selection, focusing in particular on the potential of adopting or modifying the Human Development Index (HDI) and on the problems of synthetic indices.

3.4.1 The Human Development Index (HDI)

The HDI was proposed by the United Nations Development Program (UNDP). It was developed as a composite of longevity (measured by life expectancy); literacy (usually measured in terms of a minimum number of years of schooling); and command over resources for a decent living (measured by GNP per capita). The index is based on the average relative deprivation of each country on each of these three dimensions. The similarity in content and method with the PQLI is striking.

There are obvious problems with the HDI that have been reviewed by several commentators (e.g. Lind (1992); Murray (1991)). Following the OECD approach to the analysis of social well-being, these can be grouped as:

- the choice of components and the relative weights to be attached to them;
- the analysis of how those components can best be measured; and
- whether or not the desired series can actually be measured and, if not, what new data procedures are required.

Choice of Components and Weighting

Few would dispute that the three components chosen - measures of mortality, of education, and of economic activity - are three of the most important components of the quality of life; but the choice is ultimately arbitrary. Equally, even though it may seem plausible, the choice of weights is arbitrary. These and other 'technical' problems are common to all composite indices (see Annex 3B).

Measuring the Components

Life Expectancy (at Birth)

As the measure of mortality, UNDP chose life expectancy which, in fact, summarises mortality experiences over the last 60 years. In a developing country context, indicators of child mortality tend to be more popular (e.g. UNICEF rank countries by the under-5 mortality rate in presenting data on countries in their annual publication, *The State of the World's Children*. This is because they are easier to measure in the absence of vital registration systems, and it is assumed (without much solid evidence) that all age-specific measures are highly correlated. Because of the attention being given to child survival, infant mortality is often a poor predictor of life expectancy in the developing country context (Murray 1988). The advantage of life expectancy is precisely that it does take account of mortality at all ages.

There are still problems of comparability because the implicit age-weighting differs according to the country's pattern of mortality - an economist might want to weight lives according to social output thereby priviledging certain age groups; an advocate of fundamental human rights would want equal weights; etc.. Whichever view is taken, the variability of weights in current indices seems inequitable (Murray 1991).

Lind further argues that this choice does not take into account disability which is an important consideration for industrialised societies (Lind, 1992). Indeed, in many developed countries, increasing attention is being paid to measures of Disability Free Life Expectancy rather than Life Expectancy (see ODA, 1996).

Education

The UNDP chose adult literacy because:

> Literacy is a person's first step in learning and knowledge building, so literacy figures are essential for any measurement of human development (UNDP, 1990:12).

This has also begun to be seen as a problem in developed countries, where there have recently been a series of International Adult Literacy Surveys. In most censuses and surveys, however, literacy is self-reported and not routinely assessed by enumerators (McGranahan, Pizarro and Richard, 1985). Moreover, the extent to which literacy is important is specific to culture, situation and time (OECD, 1995).

Murray (1991) further argues that investments in primary (and secondary) schooling are much more important than the attainment of literacy. While this is questionable in less developed countries, this is obviously true in 'developed' industrial societies. However, Murray's suggestion of using the educational

attainment of 15-29 year-olds in terms of years of schooling is a reversion to input indicators which is what we are trying to get away from.

Economic Activity

This is measured by GDP up to a fixed maximum ($4861 in 1994). The presumption is that above a certain floor value, individuals (or households) are no longer in poverty. This ignores the very large literature on absolute vs relative poverty and the processes of exclusion and marginalisation developing in industrialised countries. A direct measure of poverty (however defined) would appear preferable.

Moreover, even for developing countries below the ceiling, and leaving aside the problem of collecting accurate population data, it is unlikely that GDP is the best macro-economic variable for comparisons across countries.

(a) The coverage of GDP is different in terms of:

- the definition of economic territory
- the recording of taxes and subsidies
- the comparability and reliability of estimates for dwelling services (mortgages and rents)
- the comparability of estimates of the borderline between intermediate and final consumptions[20]
- the completeness and recording of the activity of financial institutions
- the quality of estimates for the transition from GDP to GNP.

( ) The extent and size of the informal sector (or parallel economy) is unknown. Both the incomes generated and the profits realised are not declared for tax purposes; however regardless of origin, the income is diffused throughout the whole economy, and this influences the level and distribution of total household disposable income.

3.4.2. Other Innovatory Composites

20 These are less relevant to developing countries; but then the problem arises of estimating Purchasing Power Parities (Kavanos and Mossialos, 1996; Murray 1991); and of GDP itself.

The use of composite indices is intended to summarise the overall impact of development changes on the quality of life for citizens of a developing country. This assumes that a statistical change can be interpreted as a real change; but there are real doubts over the validity of the data (Murray, 1991). Various other indices have been proposed as alternatives.

Johnston (l988) proposes a 'chain index' which takes its numeric values from observed year to year changes in the value of the indicators that are included in the composite index. He selects 21 statistical variables to represent prevailing conditions in nine major areas of social concern (with at least two indicators in each of the nine areas); calculates year to year percentage changes; introduces 'multipliers' to make observed variability in each of indicators roughly comparable and allowing for different weights for positive and negative movements); and adds the resulting index values to give a total score.

Lind (1993) proposes a composite index which is wider than the HDI. The HDI can record large improvements in socio-economic conditions but simply miss others. Clavijo (1992) provides an illustration for Colombia where, according to the HDI, vast improvements were recorded, ignoring the permanently high homicide rate. He goes on to propose a Right to Safety Index; and a rate of reduction of insufficiencies (RRI).

Humana (1986) has based an index on the UN Declaration of Human Rights (1948), attempting to follow the progress of these rights in 120 countries by applying various measures; and computing an overall score for the satisfaction of human rights.

Many authors have attempted to produce an index for the quality of city life: for example, the European Foundation for the Improvement of Living and Working Conditions have developed an index of the collapse of the urban environment (EFILWC, 1993).

3.4.3 Conclusion

The problems of agreeing on index components and then on their appropriate weights are extremely difficult even between a small elite group of consenting adults (for example, those sat round a central committee table). They are insurmountable if the intention is to reflect the complexity and diversity of social development, or if we are also seriously attempting to involve the community. For these reasons, a range of more pragmatic approaches are considered from 'top-down' to 'bottom-up'. In the next section, we describe the procedures suggested by Frances Stewart for developing a flexible and functioning top-down system. This is exemplified in the following two sections which assess the impact of education on health and employment. The final section considers the problems involved in designing 'participatory', 'bottom -up' scheme for monitoring and evaluation.

3.5 DEVELOPING A MODERN FRAMEWORK: MONITORING SOCIAL CONDITIONS TOP-DOWN?

Few now disagree that the impact of economic restructuring on social programmes should be considered independently of the purely economic consequences of the economic restructuring efforts. Economic development should serve human and social goals. In a statistical context, the emphasis must shift from counting products and things, to counting people, their activities, and what happens to them.

The main reason for concentrating first on the direct measurement of well-being is simply that, in recent history at least, it has been neglected relative to the total statistical efforts of governments which instead concentrate on monitoring the activities and transactions of state agencies. This does not mean that measures of the activities or outputs of government policies are irrelevant in the context of social reports; the question is to place them appropriately within a framework organised around the *intended impact of policies*, that is, the well-being of the population affected by those policies.

3.5.1 A Framework?

The problem is that the policies adopted in one sector may have unintended consequences upon conditions or trends in another. Thus, policies within the education sector have an impact upon health, the physical and social environment and the distribution of power *as well as* upon educational achievement. Vice versa, policies adopted in the health sector - sectoral policies concerned with the physical and social environment, and the existing distribution of power between various groups in society - affect purely educational objectives. What kind of instruments can be developed which take these inter-relationships into account in the process of planning and in monitoring the implementation of programmes?

The basic problem posed by this representation is to provide a framework within which information about progress towards objectives can be related to data about inputs, process and outputs of social programmes. But, as emphasised in a previous section, there are no consensually agreed and substantiated causal models linking measures of corporate or government activity (whether of input, process or output) to states of, or changes in, social welfare. The inclusion of data on corporate or government activity in such a framework, therefore, needs to be presented as the *context of*, rather than the *explanation for* the state of, or changes in, the state of social welfare; variations in the former should not be taken as *the explanation for* variations in the latter. Baster (1985) reviews many of the proposals and also cautions against:

> grandiose schemes, which attempt to capture socio-economic reality as an integrated totality, because they push more urgent and more useful exercises in statistical bridge-building between sectors out of the

limelight.

Even without making assumptions about causality, the design or maintenance of a particular statistical framework is not a neutral, technical procedure, for the way decision makers interpret current reality and conceive their future information needs itself depends upon the existing framework of data collection and presentation. The perceived appropriateness of a statistical framework for monitoring also depends on the particular social organisation. For example, capitalist states use the System of National Accounts, and what were the socialist states used to use the Marginal Product System; whether or not statistics are broken down by gender depends on the recognition of women; a land register in a desert without oil could be superfluous; and so on.

Neither can the choice and definition of what to measure and how to report it be treated as a technical issue. For example, one could argue, on practical and technical grounds, that nutrition status measurements may well be the most reliable and cost-effective source of information about social and economic conditions in countries where the health and social and statistics systems are inadequate - and that comment refers equally to many developed countries with over-developed administrative statistical systems. But the promotion of nutritional status as a major indicator is also a political judgement: thus it may be much easier to achieve a consensus in favour of using nutritional rather than socio-economic status as a basis for comparing needs, regardless of the (undoubted) value of the latter indicator.

These political influences upon the collection and interpretation of statistics must not be ignored. A framework is not a neutral, technical device: as Seers (1975) argued:

> The series a statistical office chooses to prepare and publish exercise a subtle but pervasive and lasting influence on political, social and economic development. This is why the apparently dull and minor subject of statistical policy is of crucial importance (Seers, 1975:5).

Thus, whether the aim is reportage on the outcomes of social programmes, quality of life, social progress, social trends, or in enlarging social control *á la Bauer*, the choice of what components of well-being are to be measured by indicators is essentially a political one. Obviously, any combination of these indicators via weights which would themselves be the subject of debate, would be at least as politicised.

3.5.2 Designing More Relevant Statistics?

> There are virtually no statistics anywhere on most aspects of life that really matter - the average distance people have to carry water a food; the number without shoes; the extent of overcrowding; the

> prevalence of violence; how many are unable to multiply one number by anothendr, or summarise their own country's history... (Seers, 1983:5-6)

Several sets of indicators have been proposed for monitoring development and progress over the last 25 years (for example, UNRISD, 1964; McGranahan et al, 1972). Stewart's treatment in *Adjustment with a Human Face* (henceforth AwHF) is taken as the focus here as it is both comprehensive and relatively recent.

Stewart (1987) suggested the following features of an ideal monitoring system:

- data to be available *regularly and speedily*;
- provision of indicators of the *status of the population* at large and *of vulnerable groups*, broken down by major socio-economic category and by region, to record how vulnerable groups are being affected by changing economic circumstances and by policies;
- provision of indicators of the *immediate* as well as the *underlying causes* of the changing status of the various groups;
- a well-functioning system for the *timely reporting, aggregation, analysis* and *diffusion* of the information collected (Stewart, 1987:259).

All of these are very important, as she demonstrates, and the brief discussion below - is intended only to draw attention to some of the difficulties which need to be faced before an ideal monitoring system can be developed.

Regularity and Speed

It has been recognised for many years now that the (official) statistics available to monitor human welfare status are badly deficient in most countries (OECD, 1970; McGranahan et al, 1972; UNESCO, 1979; UNRISD, 1977). This is due as much to a bureaucratic concentration on the development and maintenance of existing systems of administrative statistics as to any (intellectual or political) reluctance to consider what kind of measures would be most appropriate for monitoring human welfare status. Whatever the principal cause, however, the deficiencies of the official data are not easy to remedy (see, for example, the discussion of employment statistics in the Technical Annex).

It is worth recalling Carlson's conclusion about population data after reviewing data from 52 selected developing countries.

> "Given the limited potential of measuring infant mortality trends on an annual basis, world-wide progress in reducing infant mortality could meaningfully be reported only once every 3 or 4 years" (Carlson,1985).

In other words, while rapid feedback is important, so is precision. If we demand that a statistical system be able to monitor sensitively the fine details of changes, we are likely to overlook the biases and distortions due to inadequate data collection procedures. There is, of course, a danger in over-emphasising the difficulties of data collection and our ignorance, but too many recent publications have erred in the opposite direction (e.g. de Kadt, 1989).

Status Indicators

There should be some overall sense of demographic trends. For example, it is estimated that almost three quarters of the increase in the population aged 65 and over by the year 2025 will occur in developing areas of the world. For example, in Nigeria, the population aged 65 and over is projected to increase from 1.3 million in 1950 to 16.0 million by 2025 (more than 12 fold), and whilst the absolute increase in the USA is much larger – from 18.5 million in 1950 to 67.3 million in 2025 (Kalache, 1986) – it is *relatively* much smaller (less than 4 fold).

Stewart's brief, of course, was to suggest status indicators for children. She proposed the following list:

- Prime indicators of health and nutritional status: infant mortality rates; child death rates; indicators of nutritional status for the under-5s*; low birth weights*; indicators of morbidity (disease prevalence and incidence).

- Indicators of educational status: literacy levels; primary school completion rates; drop out rates*; repetition rates.

- Additional indicators of child welfare: rates of child labour; number of street children*.

[The asterisk indicates that the indicators will show deterioration in a short time-frame and are therefore especially relevant for monitoring adjustment] (Stewart, 1987:261-2).

It is difficult to organise the collection of accurate data in many developing countries; and it is especially difficult to rely on the accurate collection of data for many key variables in the educational, health and nutritional field. Assuming, however, that these difficulties can be overcome, Stewart's proposed list is an interesting commentary on the claim that statistical series can be fitted into a causal framework relating adjustment policies to the impact upon children, for some of her suggested indicators might be *more* appropriate for *adults*. For example, low birth weight rates are mostly used as indicators of the health and nutritional status of women of child-bearing age (Sterky and Mellander, 1973; Tanner, 1982) rather than as indicators of the child's own status; and literacy levels are, surely, more appropriate as indicators of adult status rather than of child status. In certain circumstances, therefore, adopting too definitive a framework

might not generate the appropriate policy recommendations.

Indicators which are specific to adults' well-being are:

Health and nutritional status

- height relative to average height of tallest 10% or of wealthiest group in society
- indicators of morbidity
- mothers giving birth to low birth-weight children
- adult mortality rates (adjusted for age and sex distribution)

Educational status

- literacy levels
- primary school completion levels

Economic security and welfare

- adequate resources
- numbers of dependent children

Participation in community

- crucial but only sensibly measurable at each local level. Indices have been developed for use in specific community projects (see for example. Section 2.3 above).

These status indicators, to be useful, need to be disaggregated in various ways. Rather obviously, for indicators to be useful to the local community, data have to be available on that level. More generally, the concern with the status of particular 'vulnerable' groups means that the data needed to be available for each of those groups. For example, one might want to compare for boys and girls, the levels of immunisation coverage, frequency of attendance at primary health care clinic, nutritional status, school participation rates.

It should be emphasised that, from a basic needs perspective, the issue is not relative inequality as such, but the lower and unacceptable status of certain groups.

Causal Indicators, Process and Input Indicators

Stewart suggests choosing only a small subset of indicators, those which:

- are of the greatest importance - in magnitude - in determining child status;

- may be amenable to special policies for the protection of vulnerable groups; and

- are affected by changes in the economic environment and macro adjustment policies.

The last criterion is clearly important from the perspective of macro-economists advising UNICEF; it is less clear that it is the most important in any absolute sense. The second criterion ensures policy-relevance. But the first is optimistic: we simply don't know enough to choose which indicators are most likely to reflect crucial changes.

For example, Stewart's choice of process indicators jumbles-up behavioural with environmental indicators. While the prevalence of breastfeeding and primary school enrolment rate can be treated as mainly behavioural (or at least informed by culture), the distribution of oral rehydration, immunisation level, availability of potable water and health services access rate are mostly independent of the individual. This does not make for clarity.

Her choice of input indicators is much more contentious. The focus on real incomes and real government expenditures would be almost irrelevant in some subsistence economies. Although, in earlier chapters it is acknowledged that real income data are often deficient (see Cornia, 1987:26), Stewart nonetheless recommends using other, more accessible data on basic food prices, levels of money income or expenditure in money terms, and employment and unemployment (Stewart, 1987:263).

More sensitive and useful indicators can be developed, but they depend on the particular context. While we might all agree on the kinds of things that ought to be measured, the precise definitions of indicators and the importance attached to specific data series will differ. For instance debates over the relative importance, for literacy, of the expansion of adult education or of primary schools, will be reflected in the kinds of indicator we generate.

At the same time, there are structural conditions that affect the possibility of making progress towards the attainment of basic needs. For example, there are changes in ecology and the environment, the relative accessibility of primary health care, the resource base for adult education. Some structural conditions depend upon national and international policies and provision. In addition, we must also not forget the actual and potential impact of war.

As with the status indicators, it is important to devise indicators of inequitable state inputs to services used by different groups such as the relative proportion of education expenditure on the different levels of education or the relative

proportions of health care expenditure on primary health care or on the hospital system. Similarly, we would want to compare conditions between rural and urban areas and between men and women. As MacCormak (1988) comments:

> 'a dynamic historical perspective would add weight to any analysis, but how is one to bridge the gap between inadequate, male-based statistics and informants who are very vague about the past?'

There can be no universal blueprint, for the form in which data should be aggregated in order to best serve a basic needs-oriented local development policy will depend on the precise choice of priorities made by the community. For example, different preferred household structures will lead to different kinds of data being collected and combined in different ways. Whichever data series are chosen, they should, once again, be disaggregated to the smallest possible unit.

A Well-Functioning System to Analyse and Give Feedback on the Data Collected

Stewart cites this as one of the major stumbling blocks to analysis; developing country statisticians could reasonably reply that much of their time is occupied in completing regular questionnaires from the international agencies.

Stewart is correct in pointing to the organisation of data collection and the publication of the results as one of the most difficult tasks (Stewart, 1987:264). It was emphasised at the conference on *Statistical Policy in Less-Developed Countries* held at the Institute of Development Studies in 1979 (Dasgupta and Seers, 1979); it cannot be repeated too often.

There is, however, a question as to the appropriate level of analysis and discrimination. The preference here is for this structure to be as close as possible to the community providing the data. While the attempts by UNRISD to develop a local area monitoring system have not spread, the general approach is correct. Officials record information about and from villagers and then collate the data in a form for onward transfer; they should also be able to provide a minimum interpretation back to the community, on the basis of their report, without waiting for the Ministry. Lourié (1987) also comments that the studies and surveys which do exist of the learning process and of motivation need to be published and disseminated (Lourié, 1987). They also need to be synthesised.

There are, of course, many conditions for which a local response is not adequate. Whilst a national surveillance system such as that of Botswana or Indonesia is ideal, it cannot always be replicated, or, if it could in principle, there are frequently practical obstacles. In this kind of situation, it is worth investigating the possibility of obtaining an overall view by other methods.

3.5.3 An Excess of Estimates, a Data Drought

Finally, the demand for data sometimes seems never-ending. Where, one might reasonably ask, does it stop?

It has been recognised for some time, at least by decision-makers, that there is a problem of information overload which needs to be reduced through a focus on relevant items and supporting data. This was, precisely, one of the motivations for the social indicator movement which focused on the problem of selecting a small set of data series.

The fundamental point which should be self-evident is that the collection of data is not an end in itself: it ought to be collected to answer some questions (Carr-Hill, 1987). But there are considerable pressures to extend the nature and range of data collected, including the growth of research specialisations which have meant the generation of new sets of questions (however trivial). Second, the spreading regulatory functions of the (international) state in parallel with the spreading tentacles of multilateral corporations, implies an extension of administrative activity and therefore the generation of extensive data bases.

It is, of course, true that the request for further data or further research is often a means of delaying positive social action. At the same time, it is important to recognise that the possibilities of gigantism do not obviate the need to think. Doubtful evidence proliferates while crucial data are not collected.

No-one would dispute that social conditions in many developing countries - and especially in much of Africa - are appalling and that every effort, both national and international, should be made to improve them. But global forecasts of further doom and gloom from outside, based on macro-estimates of unknown reliability, are of little help in formulating appropriate policies. In the short- and probably medium-term, it would be better to devote some statistical resources to building up a picture based on reliable evidence from local studies, rather than to continue to proliferate international 'guesstimates'.

In the long term, the development of a comprehensive national statistical framework could, in principle, provide appropriate data. But disputes over what is an appropriate pattern of growth mean that for the foreseeable future, the elements of such a system would be imposed from the centre. In the short-term, local-level monitoring with community participation is a realistic and potentially more democratic approach, and that is the subject of the final section. Before that, we consider ways of assessing the impact of education upon health and upon employment in developing countries.

3.6 THE IMPACT OF EDUCATION UPON HEALTH

3.6.1 Introduction

The purpose of this section is not to provide an extensive overview of the relationships between education and health; and in particular the extent to which education affects health[21]. For that the reader is referred to (Caldwell, 1993). Instead, the intention is to suggest indicators for monitoring the impacts of education upon health. In Chapter One, however we have already discussed the difficulty of assessing the impact of, say, the delivery of educational services upon attainment. The problem is multiplied in attempting to assess the 'added value' of education in affecting aspects of health (because of all the other factors which affect health). While it may therefore be possible to suggest indicators of the state of health upon which there is a presumed impact (and those will be considered here), it is unlikely that we can, *in general*, identify that part of any change in the indicator which is attributable to the educational intervention.

Nevertheless, many of the arguments for extending universal primary education are in terms of the improved health of school 'graduates'. No one - at least we think not - is suggesting that increased education will *per se* improve the health of the individuals involved. Instead, the presumption is that there are direct and indirect effects upon the health of an individual: direct effects are effects upon fertility, infant mortality, parenthood, attitude to positive health, preventive health; indirect effects are effects through their employability and sustainability (see section 3.7 for effects of education on employment).

3.6.2 Brief Review of Evidence of Possible Impacts

Fertility

The mission of the United Nations Family Planning and Population Association (UNFPA) is to promote the possibilities of family planning. The bulk of their activities have been through educational programmes - whether or not carried out in school.

But - unless there are large-scale tracer studies over the reproductive careers of those exposed to such programmes - the evidence of impact has to rely on estimates of changes in the population fertility rates. These are not easy to interpret.

---

21 There are many reverse influences: malnourished children are likely to drop-out, if they go to school at all (Cornia et al, 1987)); illness in the family reduces the likelihood of the children going to school (e.g. Oulai and Carr-Hill, 1994).

The World Fertility Surveys carried out by the International Statistical Institute (ISI) during the 1970s and 1980s showed large differences from indirect (census-based) estimates of either fertility or child mortality. In general the ISI estimates of mortality were substantially lower, with the most likely source of discrepancy being the misreporting of the age of mother at census.

Blacker (1987:197) concluded:

> there is still no method of estimating either fertility or mortality in developing countries lacking reliable registration which can be guaranteed to give accurate results.

Infant Mortality

The presumption of a linear relationship between female literacy and the infant mortality rate has been the basis for UN and World Bank projections of population for nearly 15 years now. We need hard longitudinal evidence to confirm this - and one possible route is through tracer studies that could be part of a monitoring system.

The evidence from micro studies is less clear. Oni (1988) reports on a household survey in Ilorin, Nigeria at the end of 1983. Out of a sample of 932 households in three residential strata, 913 married women were interviewed. He analyses for the effect of a wide range of socio-economic variables (woman's education, husband's education, type of union, religion, woman's occupation, parity, contraceptive use, area of residence, presence of indoor tap water, presence of refrigerator) on child mortality. Although all the inter-correlations are statistically significant, in a multi-variate analysis, the only 'cultural' variable that is retained is husband's education (all the other cultural variables are dropped).

Parenthood

There is substantial evidence that parents (and especially mother's education) affects child development. For example, a longitudinal case-control study in Aberdeen compared 120 low birth weight babies with 120 normal weight babies matched on gender, parity and social class (Illsley and Mitchell, 1984), and showed that the most important variable discriminating stage of development (and school achievement) at age 10 was the interviewer's assessment at birth of the competence of the mother.

It is extremely difficult to see how this could be converted into indicators of the impact of education. The 'obvious' suggestion - that one could track the number of early pregnancies - will not do, as many of the early pregnancies are associated with girls who leave school because they are pregnant.

Positive Health

One interesting sideline is the likelihood that education might lead to the collective promotion of positive health. In developed countries, an example is the socially differentiated pattern of smoking over this century. During the 1920s and 1930s, when it was fashionable, middle classes smoked more than working classes. After the Second World War, working class groups began to catch up; and then when the link between smoking and lung cancer was suggested, doctors were the first group to give up smoking, followed by middle classes. In developing countries, one can cite the (loose) links between the literacy campaigns in Andhra Pradesh and the banning of alcohol sales (Crinnion, Shotton and Carr-Hill, 1999).

Whilst these might well be some of the most important effects, it is difficult to define appropriate indicators.

Prevention of Illness

The obvious argument is in terms of the impact of education upon health behaviours; an obvious associations is the relation between education and liability to HIV/AIDS and vice versa (Oulai and Carr-Hill, 1994).

Following the Knowledge, Attitudes, Behaviours and Practice (KABP) model of health education, for example, there have been surveys of people's knowledge, attitudes and behaviours, comparing between those who have been on a literacy programme with those who have not. Carr-Hill *et al* (1991) showed in Tanzania how those who had been enrolled in literacy classes gave more correct answers to knowledge questions and appeared to have more 'modern' attitudes and practices.

We need the following:

- routine ways of measuring the extent to which health practices have been absorbed in any particular programme, whether in or out of school; and
- ways of measuring impact among out-of-school youths and adults.

3.6.3 The Measurement Problem in Evaluating Projects

The above discussion has focused on the nature of the influences of education upon different aspects of health at an individual level. When we are working at a project level, the concern is usually with assessing the benefits of any project compared with the costs. It is not possible to assign financial values to improvements of health status. In some cases, the issue is to compare the impact upon, say, rates of infant mortality, of spending on education services with, say,

spending on health services. Then one can compare the technical efficiency of the two possible projects.

This is only possible, however, when the benefits of the approach can be assessed one-dimensionally. You cannot compare projects aimed at reducing infant mortality with those aimed at reducing tuberculosis among adults.

A similar problem has arisen in developed countries, where the issue is to compare between two (or more) medical technologies and interventions. Traditionally, improvements were assessed in terms of the reduction in the death rate (or the increase in life expectancy or survivorship following an intervention). But many interventions are now aimed at improving the healthfulness of the remaining life years as well as - and, in some cases rather than - increasing life expectancy or the chances of survival. In that context, economists in the UK and the USA have been concerned to establish a 'measuring rod' for comparing the benefits of different interventions which combines quality and quantity (Weinstein and Stason, 1977; Williams, 1985). Based on questionnaires, values between 0 and 1 have been attached to a large number of descriptions of health statuses and then physicians have assigned different conditions to those descriptions so that there is now a 'quality' rating attached to patients' status before and after most common interventions. These ratings can be combined with the expected years of life to yield a Quality Adjusted Life Year (QALY); and then interventions compared in terms of their costs per QALY.

In developing countries, the World Bank have proposed a Disability Adjusted Life Year (DALY) where the indices are based on the views of 75 eminent international public health specialists. These ratings are the basis for the proposals they made in the 1993 *World Development Report*.

There are several drawbacks, many of which are summarised in ODA (1996) and are reviewed briefly in Annex 3C.

3.6.4 Conclusion

Measuring the impact of a set of services upon performance in a different domain is going to be difficult because of the (more directly relevant) service activities within the corresponding sector. While there is a large body of evidence that education does affect fertility, infant mortality, parenthood, attitudes to positive health and to prevention of illness, the problem of quantifying those impacts is made more complex because of the difficulty of defining health. The economist's proposed solution mystifies more than it illuminates.

3.7 THE IMPACT OF EDUCATION ON EMPLOYMENT IN DEVELOPING COUNTRIES

The purpose of this section is to examine briefly, current indicators used to measure the impact of education in developing countries on employment and to suggest possible areas for improvements.

3.7.1 Current Practice

How is Employment/Unemployment Currently Measured?

The labour force framework supported by the ILO and the one generally accepted world-wide, gives rise to three main categories - unemployment, employment and economic inactivity. The first two are looked at, in some detail, in Annex 3C. Since the three concepts are mutually exclusive, all those who are neither employed in some sense nor unemployed are said to be economically inactive, and this is, therefore, the residual category. Developing countries use some variation of this framework, but definitions still vary widely between them. Consequently, employment and unemployment data rarely mean the same thing from country to country.

Impact of Education on Employment

Normally, for primary and early secondary education the rate of return approach is used to assess the impact of education on labour market outcomes. As the student ages and progresses through the education system, more indicators, and more complex systems, are used to match education to the labour market so as to prevent mismatch. Thus, while rate of return (ROR) analysis is commonly used for basic education, ROR and most of the other methods described below, are used for secondary, tertiary and vocational education.

The types of labour market data that would be required by these approaches are: wages, household income, labour force status, unemployment, employment status (self-employed, employee, informal sector worker), occupational status, educational mismatch between supply and demand of qualified labour.

Rate of Return Approach (ROR)

The ROR approach calculates the net returns on educational expenditure (ILO, 1984), measured as the increase in net income that an individual will be able to command throughout his/her life in relation to the income he/she would have received if he/she had not reached that educational level.

For each specific educational programme, the present value of the flow of future net income is calculated on the basis of the above definition. Those programmes which show positive returns should be promoted, while those showing zero or

negative net present value should be reduced or possibly abandoned.

If the flow of net income is calculated as the difference between the income of the individuals after tax, and the costs to them include both the direct costs paid for the education and the indirect costs in terms of income not earned because of participation in educational programmes, for a given discount rate, this gives the *private* rate of return. If the income is calculated before payment of tax and the costs include all the resources utilised (by the individual and by the state) to implement the education programme, for a given discount rate, this gives the *social* rate of return.

According to Richards (1994), the "weapon" wheeled out to overcome the alleged negative effects of manpower forecasting on the allocation of educational resources was this "rate of return" approach. There are, however, at least five main objections to the approach. First, it neglects external effects, since the only gains quantified are those accruing to the individuals who had received the education in question. Second, the analysis cannot shed light on the extent to which households needed to be encouraged to undertake "human capital investments". Thus, for example, the persistence of primary school drop-outs co-existing with high private rates of return could be caused either by a family decision on the relative priorities of work or schooling, or by insufficient government resources to primary education. Third, the basic assumptions - that observed wages reflect the marginal product of labour, and that the content of the marginal years of schooling an individual undertakes is responsible for the marginal increase in income - are questionable. Patronage often accounts for high incomes in many developing countries, and even though the level of education might be correlated with this - rich people will send their children to schools whatever the quality of the education or of the pupil - the incomes do not reflect the education received. Fourth, it assumes that total employment remains constant. Dougherty (1985) argues that most rate-of-return studies of manpower-development programmes implicitly assume that the old post of a trained individual is not filled by an unemployed worker and that the trained individual does not displace any other worker. Hence it is implicitly assumed that total employment remains constant. Fifth, it gives no guide to the quality of education currently being given. One would have to wait at least a decade to see whether the quality of the education delivered was reflected in the wages given, which is hardly a basis for improving the quality of education today.

Several authors critical of ROR analysis, nonetheless see it as a useful technique with limitations (Bennell, 1996; Lauglo, 1996). Lauglo (1996) writes, "To give guidance for present decisions, one needs what is never available: information on future earnings associated with different types of education. Data from the past are the best we can do, and reliable estimates of lifetime income streams are only available for those educated many years ago. The problem is that labour markets and the supply of educated persons to those markets can change so as to make past income streams poor predictors of future ones."

Manpower Requirements Approach (MRA)

The Manpower Requirements Approach is another way of tailoring education to employment needs. It first came to widespread prominence in the OECD's Mediterranean Regional Project (MRP) in the early 1960s. The three major steps in manpower forecasting are: (a) projecting the demand for educated manpower; (b) projecting the supply of educated manpower; and (c) balancing supply and demand. It has fallen into disrepute because of the plethora of assumptions made and its inflexibility due to technological change.

Labour Market Information Systems (LMIS)

The seeming failure of both the RoR and the MRA to assess mismatches on the labour market has led some authors to concentrate on the preparation and organisation of labour in Labour Market Information Systems (LMIS) as an "alternative" to forecasting. (See, for example, Mason, 1979 *and* *International Labour Review*, 1994.) As Richter (1989) notes "labour market information means nothing more nor less than what it says - information about labour markets". Indeed, at best they present a shopping list of items to be collected without providing an analytical framework within which to collect and then to analyse data for planning or policy formulation.

LMIS publications are potentially useful, however, in a developing country context, to delineate the main variables of interest for manpower planning and to arrive at consistent definitions. Unfortunately most of them do not do this. Another disadvantage of LMIS, as described in the recent publications by the ILO on the subject, is that they ignore, to a large extent, the demand side of the equation. This is because data on macro-economic planning is largely the preserve of non-labour-market specialists. Since demand projections for labour depend on the economic growth rate, however, these should hardly be ignored in LMIS.

The contradiction in LMIS has, belatedly, been realised by one of its leading proponents, Lothar Richter, who notes that "the volume of labour market information produced is likely to show an upward tendency; and...as a result manpower projections of the scenario-building type... are likely to be the main beneficiaries"(!) (Richter, 1989)

Cybernetic and Pragmatic Approaches

Manpower forecasting, as we have seen, has been largely concerned to date with supply side policies and, in particular, implications for education and training. This is because the outcomes from the two main approaches to manpower forecasting, the manpower requirements approach (MRA) and the rate of return method (ROR), both concentrate on education and training policies.

The disillusionment with the two analytical tools most widely used in manpower planning has led to the development of combinations of qualitative and quantitative methods. This has been advocated by Dougherty (1985). He argues for the systematic use of all available information as feed-back for planning. Such a system should be pragmatic and eclectic through using previously neglected or non-existent types of labour-market information: vacancy/unemployment ratios, trends in relative wages, the use of key informants, etc. Such a system should be monitored by manpower planners on a continuous basis. Although he gives the name "cybernetic" to such an approach, perhaps what Dougherty really means is an "heuristic" approach to employment and manpower planning, a system to enable planners and policymakers to find out what things are necessary for it to operate effectively, given the lack of precise definition.

This heuristic approach has a number of advantages over other approaches. First, it helps to organise existing data; second it focuses attention on the labour market and the need for new research in that area; third, it focuses attention on precisely those data required to understand the labour market; and last the experiments with the system are performed in terms of scenario analysis. Thus the system avoids point estimates through giving a range of estimates that depend on a number of supply and demand scenarios.

An approach that followed a heuristic approach is the MACBETH model (Hopkins *et al* 1985). The system is embedded in a user-friendly package and results are presented in graphical form allowing a dialogue to be maintained with even the most numerically illiterate policy maker. The system is heuristic because it produces results quickly, provokes discussion on the results emanating from the scenarios and leads the inquisitive into the search for new data sources, and better ways of understanding the labour market.

Further Controversies

Among several other controversies surrounding the different approaches to measuring education's impact on employment are two important ones not yet mentioned. First, the screening hypothesis, that education merely filters individuals to appropriate occupations on the basis of their ability, rather than the education received, undermines the use of all the approaches covered above. Secondly, the old chestnut, that schooling doesn't matter, that it is, rather, social class that determines future earnings, is given some substantiation, by various studies such as Fergany (1994). Fergany finds in Egypt, that social class is a bigger determinant of future earnings, when adjusted for on-the-job experience than education per se.

3.7.2 What Improvements Could be Made in Indicators?

In their review for the World Bank s Wapenhans report, Sigurdsson and Schweitzer (1995) examined the use of performance indicators in World Bank lending for education. Little was mentioned of the relation between education and the labour

market or poverty. The report noted, however, that quality in vocational education and training (VET) was usually evaluated in terms of the employment and earnings potentials of graduates in the labour market.

What sorts of indicators could be used to measure the impact of education on the labour market and on poverty? There are two general sets of indicators: those that examine earnings and employment status today which are based on education imparted up to twenty years earlier and look at averages rather than specific individuals; and those that relate current employment or poverty status with actual education some years earlier as estimated by household surveys or tracer surveys. In each of the two cases the indicators required are likely to be similar: it is just that the methods of data collection and analysis are not entirely the same.

What improvements could be made to existing DFID practices? Judging from the Kenya study (see Chapter 2) where *no* direct account was taken of eventual labour market outcomes - admittedly this was for primary school projects where eventual poverty and labour market outcomes are far into the future - some improvements could be made. The following table suggests the types of indicators (i.e. general categories rather than specific indicators) that could be collected to measure different aspects of different levels of schooling.

Type of education	Suggested types of indicators
Primary education	earnings (wages, other earnings); rate of return (private and social); employment status (not in labour force, unemployed, underemployed, employed); poverty status (less than food-based poverty, ultra poor
Secondary education	same as primary + mismatch between job obtained and job qualified for
VET	same as secondary plus responses from employers on appropriateness of training received; indicators on satisfaction gained from work
Tertiary education	same as VET

3.8 MONITORING BASIC NEEDS AT THE LOCAL LEVEL

The purpose of this section is to examine the possibilities of monitoring distance from, and progress towards meeting basic needs at the local level. A number of basic principles are first proposed: an ideal system can be specified, quite easily, *on paper*; but that specification raises a number of issues about aggregation and about the nature of the indicators required which are discussed. Reference should also be made to the various annexes.

3.8.1 Some Basic Principles of Indicator Construction for Democratic Monitoring

First, data are *produced* not *collected*: they depend on underlying *concepts* and on a *system* of processing in which different agents have different interests and tasks. Equally, the historical and social context of measurement is important: for example, the ready access of quantitative measures and techniques for aggregates of things has dominated the way in which statistical systems have developed. In sum, measurement work and statistical work are not socially or theoretically autonomous activities.

Consequently, the activity of measurement itself is a potential agent for change. Indeed, the potential of data measurement to influence policy often leads to its suppression, even when no-one disagrees about the concepts or definitions. We also have to recognise the political role of information. For example, educational and health professionals do not necessarily use data to present their case to the

right people. This is graphically illustrated by Gordon (1979), who relates her experiences as Director of the Bakwu Applied Nutrition Programme in Ghana over a five-year period.

Second, everyone might agree that a particular phenomenon is worth measuring, but the actual indicator chosen would vary according to the clientele. Consider, for example, school attendance, which everyone wants to know something about. The government planner, typically, will be interested in enrolment, repeater and drop-out ratios, pupil-teacher ratios, construction costs and so on; people would be more interested in access to different types of educational facilities, what they or their children can learn in different institutional contexts (it need not, of course, be a building, or even a formal programme); and concerned pedagogues in the type of resources that are needed to impart the type of knowledge which is socially useful.

Third, the same indicator can be *used* in various ways. Thus, an indicator of individual well-being may reflect a current *condition*, membership of a *risk* group, or a *trend* in the causative factors. Accordingly,

> a change in the use of an indicator from, for example, the diagnosis and treatment of malnutrition in the individual, to the quantifying of risk for families or communities, or to the analysis of trends and changes, requires a change in definition and significance of that indicator. This dependence raises fundamental questions about the procedure for defining indicators, about who should be involved in the process, and about the role and objectives of research (Dowler *et al*, 1982:101-102).

In general one must be very wary of how an index is used, as opposed to how it was developed.

Finally, since social change can only be carried out by people, measures and statistical activities should be on the human level and, as far as possible, organised around their possibilities for change. In principle, this means that we have to understand how people develop their own goals in their social environment and how they develop their own measurement criteria. In practical terms, many authors have remarked that the validity of data depends upon the extent to which the informant understands and agrees with the motivations and objectives of those collecting the data, and at least consents to the use to which the data will be put. Even this pragmatic approach imposes severe constraints on the viability of surveys which are centrally designed and executed.

A corollary is that measures and statistical procedures should be transparent. There is a need for a middle way between everyone having to become a statistician overnight, and the statistician's responsibility for explaining the assumptions as well as the implications which varied assumptions would have. The obvious example here, is the ease with which an economic statistician slips from talking about economic welfare to measuring GNP *per capita*, without

explaining the limitations of using the latter as a proxy for the former.

3.8.2 An Appropriate Local Information Monitoring System

The basic proposal is not, in principle, complex, assuming agreement at a political level about what are minimum standards. It would include the following:

- Agreement at a political level about what are minimum standards and about statistical specification of these standards. The same people, of course, might be involved in both the political debate and the technical development.

- An accurate picture of the present situation. This is not easy, but the information needs to be presented in a form that is digestible but not pre-digested (already interpreted according to a particular schema). Unfortunately, present statistical skills are not oriented towards the task of presentation: it is not only a task of instigating numeracy among the non-numerate; it is equally important for the data manipulator to learn how to communicate.

- A documentation of appropriate resources. This is by far the most difficult because of the lack of substantiated theory as to what kind of resources (of land, labour and capital) are most effective with different social organisations in attaining which levels of welfare for individuals. We have a general idea as to which resources are necessary, but not those which are sufficient to attain desired ends.

Overall, the statistical system would resemble more an inventory of opportunities for activities and of possibilities for the attainment of basic needs, than a recording of current stocks and flows of products and services (for two interesting exercises see Annex 3D).

The Nature of the Data

The most appropriate form of data collection will vary. Thus, finding out what are people's objectives is, at best, organised collectively. Relevant background data would be made available, and new data might be required; but the appropriate method cannot be decided in advance. Even in a less than ideal system, we would envisage that informal appraisal, inspection, or the interviewing of key persons would be more appropriate than an attitude survey.

There are two difficulties with the traditional survey approach: first, the answers to questions of the form 'which of the following do you think are the most important?' when asked by a project team with a specific interest, are almost certainly biased. Faniran (1986) reported on a survey of rural households perceptions of water quality in three rural communities around Ibadan, by a water project team, that people gave highest priority to the provision of potable water among their preferred

amenities, and 80% gave water as first and second choice. Another project team working in a different sector, would have obtained a very different answer.

Second, the answers to general questions about objectives and political strategies for attaining them are open to a wide variety of interpretations. Consider, for example, the assessment of whether or not communities really have control over the decisions which affect their daily life. Although their perceptions may not, of course, tally with those of the researcher, people usually do have quite clear perceptions about the extent to which they have control over their daily lives. But they are conditioned, and sometimes informed, by the external constraints upon them and, despite the sophistication of the various scales which have been developed by social psychologists purporting to measure autonomy, control and fatalism, these are related to the individual and not to their social context nor to the collectivity which is the issue here. Moreover, there is no obvious way in which communities can express their lack of control other than by affirming it - often in very banal ways (although a modest proposal is made in Annex 3B).

We can be more specific with the next two stages. An assessment of the present situation requires a household survey carefully designed at the local level. An informal appraisal could be wildly erroneous, and administrative data are likely to be systematically biased. Finally, an inventory of existing and potential resources would basically rely on collective discussion at a local level supplemented by household survey data. The only role for administrative records would be in providing background data on the existing situation and for an inventory of possibilities in respect of the more macro objectives, such as the equalisation of disparities between population groups.

Much of the data required by such a statistical system need not be very sophisticated. Far too much effort is often invested in both developing and developed countries to obtain precise numerical information about a state of affairs where nominal data would do. For example, it is interesting to know how well someone can cook, but what is important is that we will not be poisoned by what we eat. Similarly, the well-known tendency for the qualifications required for a given occupation to be upgraded over time (Dore, 1976) is often accompanied by an increasing degree of differentiation in the level of qualifications obtained and therefore in the complexity of administrative records. Yet, all we really want to know is whether the person can do the job or not.

The problem of aggregation

When such a system is spelt out in detail, it will appear sensible to aggregate and compare according to some common denominator (whether in terms of resources such as labour or capital, or at an intermediate level in terms of particular levels of goods and services). The mistake which has been made all too often is to aggregate too far, too fast. For, although the physical quantities (whether of land, labour and capital, or of goods and services) *can* be converted into a common monetary or other unit and then aggregated, *this* conversion *assumes* a whole

structure of production, distribution and exchange. It is therefore completely illegitimate to suppose that this aggregation has any meaning outside that particular structure. A more homely (North) example might help here. With fiscal harmonisation of the EEC, the problem of harmonising government subsidies to physical and social infrastructure has arisen. In turn, this requires ways of assessing the quality of services that are provided. How do you trade off, for example, the exceptional convenience of the Paris Metro as compared to the London Underground against the (relative) lack of violence on the latter? The answer is that those kinds of issues have to be decided not by the parachuted-in-social scientist, but as part of a democratic political process.

Obviously, it is sensible to bring together and compare findings from surveys in different areas (Edingbola et al, 1986), but it does not follow that the results should be aggregated to present an overall picture. For the particular mode of aggregation - including simple addition - will benefit some groups and disadvantage others. Compromises are made in the process of decision making, depending on the relative power of different groups. For a *post-hoc* analyst, these compromises imply a mode of aggregation and a set of weights; but they cannot be taken as the appropriate basis for future decision-making. The pressure to substitute a technical form of appraisal (e.g. cost-benefit analysis) for the essentially political process of decision making must be resisted.

Utopian - or An Essential First Step?

Obviously, one could criticise these suggestions as being impossibilist and utopian. But the experience with the failures of development planning where mistakes are made simply because we do not know very basic facts about the target populations - or because what we do think we know is very partial - suggests that ensuring that there is a firm base for knowledge is vital not a luxury. When this is allied to substantive democratisation and community based participation, then there have to be some moves in the direction suggested.

Of course, there have to be caveats about the tendency for local élites to control and thereby generate more inequity; but that is true of all attempts at localisation. While we do not know the answers, we are at least aware of the right questions to ask: thus it is important to know **who** decides **which** kind of information is to be collected and for **what** purpose. Nevertheless, the argument here is that we cannot know the true extent and nature of poverty (or progress) unless those immediately concerned are also involved in measuring and monitoring their own status.

ANNEX 3A

Opportunity Cost and Valuations

The purpose of this appendix is to review briefly the development of synthetic indicators based on the National Accounts and those developed by using valuation methods based on opportunity cost, whether measured in terms of monetary values, or in terms of time.

Extending GNP: Modifying the National Accounts

National Income can be regarded as a kind of index which (in principle) adds together all goods and services produced in an economy, using as weights either their market prices or the cost of inputs used to produce them (their factor cost). Although not originally intended as a measure of welfare, even economic welfare, it is often used in that way, in spite of many criticisms.

In attempting to avoid the arbitrary choice of weights associated with the HDI type of index (see Chapter Three, section 3.1.1.), one approach is to modify National Income so that it better reflects welfare. This can entail including non-marketed production and other "goods" such as leisure, making deductions for production which does not contribute to welfare or for social or environmental costs, reclassifying items among consumption and investment, or among intermediate and final production, and so on.

This approach requires that all the goods and bads introduced into the modified National Income be assigned a monetary value. Where, as in most cases, they are not traded in markets, some indirect valuation method has to be used. The development of cost-benefit analysis and, more recently, increasing concern with integrating environmental costs into policy making has led to the development of a number of valuation methods, mostly based on the "willingness to pay" principle.

Attempts at Adjusting National Income

The development of national accounting was marked by a number of debates, most notably between neo-classical economists interested in arriving at a measure of economic welfare and Keynesians concerned to create the tools required for demand management (Seers, 1975). The consensus which developed around Keynesian policies after the second world war meant that the Keynesian view was the one embodied in the national accounting system adopted by most governments. National Income was not intended as an index of welfare (or indeed as a measure of income).

Nordhaus and Tobin (1973) were the first to propose and estimate a modified version of National Income, intended more fully to reflect economic welfare. They proposed three kinds of modification. First, there was a reclassification of expenditures, with health care and education treated as investment in human capital, and certain expenditures, such as on the police and on defence, treated as

"intermediate", that is, not in themselves generating welfare. Second, imputations were introduced for the services of capital goods such as owner-occupied dwellings and durable consumer goods, for leisure time, and for some forms of non-market production. Third, some costs of urbanisation were deducted.

While maintaining that GNP is deficient as a measure of welfare, Nordhaus and Tobin argued that it is sufficiently well correlated as to make separate measurements unnecessary. This view is disputed by Daly and Cobb (1990), however, who argue that shorter time periods than the 1929-65 period examined by Nordhaus and Tobin, reduce this correlation (Daly and Cobb, 1990:76-80). Further, such a correlation is a likely consequence of the imputation methods used.

Other versions of measuring national welfare were implemented, such as the NNW for Japan (already discussed in Chapter 3, section 2.1). Another measure was constructed by Zolotas (1981), an Index of the Economic Aspects of Welfare, which took into account pollution costs and resource depletion.

The most recent attempt at a modified National Income is Daly and Cobb's Index of Sustainable Economic Welfare (Daly and Cobb, 1990). This is an attempt to draw on the best of previous attempts, while improving on their limited treatment of environmental issues and sustainability. The ISEW starts from personal consumption, which is weighted by an index of distributional inequality. Based on a search to quantify the annual welfare flow, and excluding the stock of capital welfare, a number of items are added: household labour, services derived from consumer durables and highways, public spending on health and education, and net capital growth. Other items subtracted include: spending on consumer durables, health and education, advertising, various urbanisation costs, and pollution and resource depletion costs.

The Proposed UN Satellite Accounts

Another approach to modifying the National Accounts was through the construction of satellite accounts linked to the main national income accounts, as discussed among international organisations for at least twenty years. This discussion has coalesced into proposals for a "Satellite System for Integrated Environmental and Economic Accounting" (SEEA), described by Bartelmus et al. (1991).

Satellite accounts represent a compromise: definitions, accounting identities, production boundaries, etc. are consistent with the core accounts, so that the two can be combined, or not, according to intended use.

The focus of the SEEA is narrower than that of the other welfare measures discussed above, since it is concerned only with environmental issues. If and when implemented, however, it will provide a measure of environmentally adjusted National Income that, in principle, is a better index of welfare than national income itself. The development of other satellite accounts (incorporating for example other social costs of production), while at least as difficult, would make it possible to

arrive at a better index.

Evaluation of Using the National Accounts

There are a number of problems faced by attempts to base a welfare index on modifying the national accounts, some common to welfare indices in general (which are considered below), but most of them to do with the difficulties of monetary valuation. We attempt here a **pragmatic** evaluation.

Familiarity with aggregates such as GNP, and with its use as an (imperfect) welfare indicator, obscures the **full extent** of its shortcomings. Hueting (1991), for example, discusses some fifteen objections to National Income as a measure of welfare. Many of these objections have been part of the debate about GNP since the 1960s and are addressed, at least to some extent, by the indices discussed in Chapter Three, section 2. They include expenditures currently classified as final but which should be treated as intermediate, a major part of government expenditure, as well as all the aspects of welfare which are not measured by human production of goods and services (environmental goods, inequality, leisure, working conditions, security, etc.)

But some of the objections (although described by Hueting as being of a "technical" nature) are in a sense more fundamental, because they undermine the basis for using market prices at all: national income omits the consumer surplus, although this is part of welfare; it involves adding different people's utilities, which is unjustified particularly where there is inequality of income; it ignores diminishing marginal utility, of individuals and of the economy as a whole; real national income requires the use of price indices which can only be calculated correctly for a constant basket of goods; and the movement of activities from the unpaid to the paid sector is ignored.

Other problems arise in relation to the policy relevance (or otherwise) of the national accounts. Seers (1976) emphasises, for instance, the "monistic" character of the accounts, with their treatment of the whole nation as an appropriate object of analysis and policy. In a period characterised by halting GNP growth, combined with increased inequality and poverty, restructuring of production, globalisation, and environmental problems which cut across national boundaries, this national emphasis is unjustified. It is more important to consider how the accounts should be extended to monitor developments in the international economy on the one hand, and with breakdowns by income classes and by production sectors on the other. This applies to the current national accounts, and equally to attempts to create welfare indices out of them.

Imputing Monetary Values

There is now a large literature (barely touched on here) concerned with imputing monetary values to goods and bads for which there is no market and thus no market price, with particular emphasis on environmental costs and benefits. The

methods proposed are mostly based on the "willingness to pay" (WTP) principle (or its twin, "willingness to accept compensation"), which rests on the assumption of "rational" economic people, spending their incomes in such a way as to maximise their welfare. In the case of marketed goods, the observed willingness to pay 1 Euro for a good is taken as evidence that the good procures the buyer at least 1 Euro's worth of welfare. Summing over goods and individuals then gives a measure of welfare, and this is essentially the rationale for treating National Income (NI) as such a measure (although it ignores several problems discussed above, and others).

Deriving Estimates for Willingness to Pay (WTP)

Where goods are not marketed, WTP is not directly observable and can only be estimated indirectly. One popular approach to environmental valuation, for instance, is to design an econometric model which incorporates all the factors affecting price, of, say, housing, and an estimate is obtained of the variation in price - and hence WTP - which results from differences in the amount of an environmental cost or benefit. Particular problems arise from the need to include **all** factors affecting price (hence there are huge data requirements), and from the fact that in practice environmental factors are highly correlated. At best the method can be used to estimate WTP for **some** aspects of **some** environmental goods.

A more experimental approach, contingent valuation, consists in simply asking a sample of people about their WTP, either directly or through a more elaborate game playing strategy. This method is cheap and generally applicable; but there is a major problem of realism, and substantial observed differences between willingness to pay and to accept compensation. In the latter case people may give infinite values.

There are many other problems with WTP indices, not least of which is their close relation in practice to **ability** to pay. Thus, utilising an index based on WTP as a basis for policy targeting is likely effectively to assume that poorer countries, regions and people are "underpolluted" relative to richer ones.

To arrive at an index of welfare, WTP methods have to be applied not just to environmental factors, but also to social and economic factors operating outside the market. This raises a whole host of problems of definition and interpretation.

Problems of Definition and Interpretation

Before valuation methods can be applied to the various costs and benefits relevant to welfare but presently omitted from national income, these have to be agreed on. There are problems of definition, of what to include, and of uncertainty. As an example, "defensive expenditures" - spending designed to offset environmental or social degradation - is currently included in national income, but should be deducted. Here valuation is not a problem, but agreement on which expenditures lead to a genuine increase in welfare and which expenditures offset declining welfare is difficult to reach.

Uncertainty raises further difficulties, particularly in relation to environmental costs. The likely future costs of global warming, for example, are widely thought to be very high, but there is great uncertainty about how high, what they will be, and how they will be distributed (e.g. among nations). Simply omitting costs we are uncertain about may lead to serious overestimation of welfare. Introducing these costs as knowledge of them improves will create problems of stability and comparability.

Valuation Methods Based on Time

The same kinds of methods are being applied to combine other sets of indices, one example being the development of time budgets. Thus, modifications of National Income focus on valuations of economic activity. Another approach, taking the same 'utilitarian' framework, starts out with the 24 hours in a day as the fundamental unit.

Much of society and social theory defines people in terms of what they do **as paid employment**, and many social scientists and particularly psychologists regard paid employment as a key determinant of individuals' well-being - or lack of it. To an extent this is understandable since what a person does to earn a living is often the second largest **single** occupier of their time every week (sleep being the largest).

The reality of work is, of course, far more complicated: many jobs require a variety of different manual or intellectual skills; some are physically exhausting but require little or no intellectual or emotional input; others have no apparent worth for the individual worker beyond a wage-packet at the end of the week. Since it is also true that about half of the population does **not** do paid work in a formal sense, it is clear that to obtain a full picture of human activities and well-being in our society we must broaden our perspective.

In recent decades, and particularly in the light of the work of the feminist movement there has been a marked increase in the recognition of the importance of housework and childcare as significant occupiers of time for many people. Perspectives, and indeed practices, have significantly changed in the last three or four decades, particularly with regard to the division of unpaid labour between men and women, the status of childcare in terms of work and leisure, and the benefits and drawbacks of technological innovation with regard to unpaid work.

Aside from these compulsory occupiers of time (i.e. 'work' in its broadest sense) there are also changes occurring in 'free' or 'leisure' time. Social commentators have begun to take an increasing interest in 'leisure' whether as a source of emancipation and self-directed activity or of passive boredom.

Time Trade-Off Methods

Richard Stone developed a System of Social and Demographic Statistics based on the length of time spent in any one status; that approach was discussed briefly

in Chapter 3, section 2.1. There have been some recent attempts along similar lines discussed below.

Quality Adjusted Life Years

Disability Adjusted Life Years (DALYs), and their functional equivalents such as Healthy Life Years (HLYs), for example, are calculated by subtracting from life expectancy the average length of time spent incapacitated (or seriously ill). However, as in practice in Japan and the United States, they involved only marginal adjustments to life expectancy itself, little further effort was made to develop them.

Using Time as the Numeraire for Modifying GNP

A totally different approach has been suggested by Lind (1993) as an alternative formulation to the Human Development Index. The argument is that instead of a simple average of relative deprivation, the best way of combining life expectancy and GDP is with the compound:

$$L = b^{w} * e^{1-w}$$

in which 'b' is a measure of income, 'e' is life expectancy, and 'w' is the proportion of time spent in employment.

However, this formulation makes the implicit presumption that the quality of non-work time is independent of income - which is a little doubtful.

Collecting Time Budget Data

Time budgets consist in recording at fixed time intervals what activities a sample of individuals are engaged in, so that the proportion of people engaged in different activities at any one time, or the length of time spent on different activities, can be estimated. Since all activities take place over time, time budgets provide a framework for considering activities systematically, although of course the welfare associated with activities cannot be reduced to the time (any more than the money) spent on them. Time budgets allow us to assess the importance of a broad range of human activities: anything that occupies a significant amount of one's time is likely to be of personal psychological and social importance.

Time budgets also provide an opportunity to get away from the emphasis on employment time, which is a feature of most existing statistics, as well as the dichotomy between "work" and "leisure", which increasingly fails to capture the reality of how most people spend their time.

The available data, however, draw a clear distinction between paid employment and other activities. Thus while information about the conditions of paid employment has been of increasing interest to government for a century, statistics about other activities are comparatively scattered and few.

An Index of Economic Welfare

Three approaches have been reviewed: extensions to the national accounts; imputing monetary valuations; and using time budgets.

It should be emphasised that, although there has been considerable inter-governmental discussion, the SEEA (Satellite System for Integrated Environmental and Economic Accounting) is only a proposal: the other measures are accompanied by estimates, however heroic the assumptions on which they are based. While many of the data required by SEEA have been the subject of research studies, no country comes near to having the range of estimates required. Bartelmus *et al* illustrate the proposals with data from "a realistic, but fictitious country" (1991:140).

Moreover, as Daly (1989) emphasises, the purpose of the reform of the SNA to include satellite accounts is in order to create a better measure of **income**, making it clear that much else needs to be done to arrive at a measure of **welfare.**

In practice, even the best attempts so far at indices of welfare have been forced into arbitrary assumptions. For example Zolotas (1981) assumes that exactly half advertising is of a "persuasive" kind (and thus should be excluded). Daly and Cobb assume that half higher education spending is consumption and half investment. There are many other examples.

Attempts at monetary valuation of non-marketed goods have to make a lot of assumptions about how people actually reach decisions which are not borne out by empirical data. Reaching valuations usually involves ignoring 'irrational' responses, and/or averaging across a very wide range.

On the other hand, as we have seen, there is often a high degree of correlation between the conventional aggregates (NI, GNP) and the modified versions. This results partly from their common core (most marketed production). In addition, large items such as the value of housework or of leisure are estimated essentially by multiplying the time involved by the average wage rate or something closely correlated with it. The latter is likely to change much more quickly than the former, and is of course highly correlated with NI. Yet the pervasive use of money and prices as a method of valuing welfare may becoming less and less relevant.

Using time budgets is, of course, completely different; but it also relies on a single method of valuation. It is this presumption - that there is a single method of valuation that can be applied to the whole range of human welfare - which is most questionable, and was one of the main reasons for developing social indicator systems in the first place. Finally, none of the proposals tackle any of the problems of monetary valuation or of relying on a single index discussed in Chapter Three.

ANNEX 3B

Technical Problems in Developing Indicators

> "The series a statistical office chooses to prepare and publish exercise a subtle and pervasive influence on political, social and economic development. That is why the apparently dull and minor subject of statistical policy is of crucial importance' (Seers, 1975:3)

Although the title of the appendix is 'technical' problems, the concern here is to identify the policy implications of some apparently technical choices. In particular, we should repeat that the concern is with measuring outcomes rather than inputs or activities. It is therefore, absolutely crucial to establish what we are trying to measure before elaborating statistical procedures for collecting data for indicators.

Three issues will be dealt with: first the pitfalls involved in deriving indicators: second, the difficulties of constructing composite indices; and third, the special difficulties of monitoring the disadvantaged.

Deriving Indicators

> There are several technical problems. We consider just four: objective and subjective data; issues of comparability; reliance on existing data; formulating indicators.

Objective and Subjective Data

There is a general failure to incorporate both 'objective' and 'subjective' components into indices; or to know how to combine them. This is partly because of an over-rigid division between 'qualitative' and 'quantitative' methods of collecting data; and partly a failure to recognise that much of human behaviour is governed mainly by how a situation is perceived, not by objective circumstances (e.g. Keynes' analysis of the behaviour of markets; Thomas' analysis of the behaviour of Polish peasants).

For example, a concern with 'exclusion' - although resulting from objective circumstances - is, at least in part, a subjective matter in that the criteria which identify the excluded are socially determined. There is a need for both quantitative and qualitative data, and they must be open to revision and disaggregation.

The usual difficulty with subjective data is that respondents are being asked their opinions about questions framed by the policy-maker/researcher. While there are other possible approaches, they are time-consuming.

Comparability

The attempt to derive a set of social indicators or a societal index which is relevant across all developing countries raises issues of comparability and the possibility of comparative research. Living conditions can only be monitored and compared if the macro frameworks are similar, including 'the nature of economic and social restructuring, the role of the state at a central and local level, the relationship of bureaucratic to market allocation mechanisms, the ideologies which underpin the social institutions of family, religion, education and so on.' (Harloe and Martens,1984). They go on to argue that, at the micro level, the situation within each country or region 'can only be understood within the context of the particular legal and financial arrangements, financial structures, cultural and religious traditions and geographical settings'. While, as a consequence, 'each observation is unique and generalisation impossible', at the higher level, 'common cross national trends may be identified even if they manifest themselves in very different forms in different countries'.

This has consequences for the kinds of social indicators systems that should be developed. For example, a cross-national concern with the 'informal sector', unless specified more precisely - may be translated differently in different contexts.

Reliance on Existing Data

The content of most quality of life measures is dictated by data and measures that are readily available rather than the demands of prior theory. Etzioni and Lehman(1967), at the beginning of the 'social indicators movement', pointed to the problems this can cause, for example:

- 'fractional measurement', where there is a lack of correspondence between a concept and its operational definition (as with 'unemployment' and its measurement using social security statistics);

- 'indirect measurement' (e.g. measuring educational attainment by years of school), especially prevalent where data is used which was collected for some other purpose.

Worse still, many measures were in fact established on the basis of a previous administrative or, in some cases, theoretical framework. There are two specific concerns here, which have been mentioned earlier: the tendency for data series which are related to the national accounts to have been developed more systematically than others; and the pressure to use the existing investments in administrative data systems to derive social indicators.

Formulating Indicators

In many or most cases the main concern is not with averages but with the proportion of people who achieve an acceptable minimum standard - that is, with

avoiding poverty and marginalisation. This concern moreover is not simply with ensuring a basic income for everyone; it cuts across all the other concerns and includes, for example, basic literacy, avoiding early death, and the availability of transport providing access to basic facilities.

Agreeing on minima within any one cultural group is usually possible but reaching agreement on such minima across groups or societies - other than for the pre-requisites for survival - appears to be very difficult; whilst attempts to define a level relative to each society (e.g. the bottom 10% of that group's or that society's income distribution as poor) are rather vacuous. Minima have to be context specific.

It is also important to establish an "optimal need satisfaction", beyond which level, more is not necessarily better (eg. food, housing, etc.). Indeed more can even mean worse not only in physical terms (eg. vitamins A and D in excess) but also in terms of excessive security and overpowering relationships. This means that in many cases, neither a set of 'positive' indicators ('the proportion who are literate') nor a set of purely 'negative' indicators ('the proportion without adequate housing') will be sufficient: instead, it implies relatively sensitive measurement of the proportions who are substantially below and substantially above that group's **critical optimum**.

It is also important to point out that in many or most cases the main concern is not with averages but with the proportion of people who achieve an acceptable minimum standard - that is, with avoiding poverty and marginalisation. This concern moreover is not simply with ensuring a basic income for everyone; it cuts across all the other concerns and includes, for example, basic literacy, avoiding early death, and the availability of transport providing access to basic facilities.

Difficulties with Composite Indices

There are many difficulties with devising an overall quality of life index. In fact, there are two distinct sets of problems: establishing a coherent set of component indicators; and interpreting combinations.

Choice of Components

There is no consensus over what should be the components or the weighting procedures that should be employed in 'composite' quality of life indices:

- whilst everyone needs/wants a certain minima of several conditions, few can agree on what is the optimum level or what combination is required;

- whilst nodding in the direction of consumer sovereignty as the mechanism for choosing components and how to combine them, few have actually attempted to take that position seriously.

There is the counter argument that each of the components is the product of a gradual developmental process out of which some degree of consensus has emerged. But that argument also is the foundation for the objection that it is an *historical* consensus. Whether or not such components or weighting are relevant or 'salient' to different population groups today is important. Whilst there is equally no consensus as to how relevance or salience ought to be measured, nor differences reconciled, if public perceptions are to be an eventual component of their experienced quality of life, then the relative importance of different aspects of their situation must also be essential.

Trade-Offs are Obscured

Although there is a very close correlation between life expectancy and per capita income (also of PQLI and per capita income), the relationship of economic growth and development is not so simple e.g. isolation of elderly and certain forms of child abuse prevail more in high income nations.

For example, Etzioni and Lehman (1967) argued against 'formalistic-aggregative measurement of collective attributes', as with the US crime index, which simply adds up a broad range of crimes, giving the same weight to a murder and a $50 theft. Seers (1985) goes so far as to suggest that synthetic indexes are often constructed without understanding the system that generates them.

The emphasis is on individuals, looking at extremes rather than averages, and not lumping them all in together for a set of bland results.

The point is that not only is well-being multi-dimensional, its aspects are incommensurable in that although they are inter-related, they are not substitutable for each other. For example, a sufficient income to ensure good nutrition increases life expectancy, but you cannot compensate early deaths with high income. **Indeed, although an index, through continued use, can be presented as being simple - as has, one might argue, GNP itself - the underlying presumptions are often quite complex and obscured.**

Lack of Disaggregation

Few quality of life indices address distributional aspects of the different components of the 'quality of life' or 'well-being' of particular population groups.

This is principally because of the difficulty of collecting sufficient nationally comparable data to yield meaningful estimates at the community level or for small groups; such indexes can usually only be calculated for highly aggregated and often inappropriate geographic units of analysis.[22]

---

[22] This is also sometimes because of confidentiality: For example, in the UK where data at small area levels or for small groups are 'Barnardised' - that is -1, 0, or +1 are randomly added to the counts.

Monitoring the Marginalised

There are many practical difficulties with assembling data whether we are talking about administrative systems or about social surveys. But there is a general difficulty which should be mentioned, given:

- increasing - or increasingly recognised (because of the 'new social movements') - diversity within the population;

- the purpose of many of these systems is to monitor the living standards/ quality of life of those at the bottom of the heap or who are marginalised; and

- our overall concern with empowerment - or providing information important and useful to the concerned citizen.

This general difficulty is that the categories used in administrative systems and in most social surveys are often contested by those being monitored. Problems of measurement then arise since, while participation of the subject community under study is always vital, the strongest groups tend to be the most vocal and visible, and their perceptions of the appropriate categories may prevail

More Sensitive Social Surveys

The problems of 'monitoring the disadvantage' is only an acute form of the general problem of monitoring the quality of life where there is increasing attention being paid to user orientations.

One possible proxy approach, in general, increasingly used in UK and USA is that adopted in the survey conducted for the ITV in 1989, published as *Poor Britain*. Respondents were asked to select from among a long list those items they thought were essential/necessary for civilised living in today's Britain. They were then asked whether they themselves had access to those items. This approach has been adopted in developing part of the study design and questionnaires to evaluate the impact of devaluation of the CFA upon poverty in the Ivory Coast (Carr-Hill, 1995).

An extension of that approach - which has been tried within the health care context - is to ask people to name which are the most important sub-components of any area of concern; and then to ask them how they rate their own situation or status in respect of each of these components (Ruta *et al*, 1994). Whilst this methodology is still in its infancy and, as currently designed, tends to favour the articulate middle class respondent who is used to reflecting upon and rating options, it might be worth exploring.

Whichever avenue is adopted, the importance of focusing on user orientations and views about what constitutes the quality of life should not be ignored.

A Local Approach

A local approach to developing quality of life indices should include:

(1) observations on the locality ranging from total area and population, to number of post and telegraph offices to number of tea shops, to prices of goods and clothing;

(2) listing of all households in the locality (c. 2000) with standard socio-demographic information; and

(3) an omnibus survey of a sample of 100 households in each locality (consumption, education, health, etc.).

ANNEX 3C

VALUING HEALTH AND MEASURING (UN)EMPLOYMENT

Valuing Health

Theoretical

People do not agree on the definition of health. Even in the basic questionnaires, doctors and patients have different views (Kind and Rosser, 1977). The basic problem is that people have different definitions of health. Blaxter (1981) has identified at least five dimensions. Even if one agrees that, in principle, this is a worthwhile exercise, in practise this makes valuation very difficult.

Should we value patients or groups of patients equally?
In fact, both the QALY and the DALY give decreasing weights monotonically as people age (because they have lower life expectancies). However people are not so simple: there is survey evidence that the young and old alike place more weight on parents with young children and less on the old and very young (Wright, 1985)

Does the same value of the index mean the same thing to different people?
People who suffer from illness generally adapt; and even if they do not adapt, their perceptions of health and illness may change. Some people may be more independent than others; etc.

Despite being sometimes thought of as a utility measure, the QALY or DALY is at best a pseudo-utility measure, (Culyer, 1990) because it does not include any consideration of non-health welfare.

Practical

The way in which the index is constructed means that quality scores tend to be compressed towards one; so that even though health deteriorates with age, older people when asked will still claim to be satisfied with their health (Wright 1985).

The approach uses a discount rate - the DALY was discounted at 3% - but there is little survey evidence to support this. High discount rates lead to a bias against educational or other interventions at young ages.

Who should decide on the quality scores? The QALY procedure is supposed to be 'democratic' because people are asked to rate health status: in fact they are entirely constrained by the design of the questionnaire instruments (Carr-Hill, 1992).

Planning Context

Given the move towards public participation, is it appropriate to introduce an index as a basis for decision-making where only a limited numbers of 'experts' are - or could be - conversant with the criticisms?[23]

The costs per QALY or cost per DALY figures are based on average costs. Such figures may not be optimal; and may not be easily transferred from one context to another (let alone from one country to another). Marginal costs of an additional intervention may be different from average costs.

The approach may give you some idea of what to aim for - the best package of services - but no idea how to get there.

Employment/Unemployment Concepts

Unemployment

The conventional definition, and the one most widely used today, was agreed upon by labour statisticians at the 1982 ILO conference on the subject. Briefly, it sets three separate criteria for classification as unemployed. The unemployed must first be without work; second, currently available for work during the reference period; and third seeking work, that is, they must have taken specific steps in quest of a job during a specified recent period.

Thus, the unemployed comprise all persons above the age specified for measuring the economically active population who were: (a) "without work' (and not self-employed); (b) "currently available for work"; and (c) "seeking work".

Special provisions are made for persons without work who have made arrangements to start work at a date subsequent to the reference period and for persons whose employment contract is temporarily suspended.

The "without work" criterion draws attention between employment and non-employment. "Without work" should be interpreted as total lack of work, or, more precisely, as not having been employed during the reference period. The purpose of the "without work" criteria is to ensure that employment and unemployment are mutually exclusive. A person is classifiable as unemployed only if it has already been established that he or she is not employed. Thus persons who were engaged in some casual work while seeking employment should be classified as employed, in spite of the job search activity. The other two criteria of the standard

---

[23] Of course, one cannot expect everyone to be conversant with the technical calculations and procedures underlying many such indices; but we are referring here to the issues of valuation (discussed under the theoretical heading above), which underpin the development of such indices.

definition of unemployment, "current availability for work" and "seeking work", serve to distinguish those of the non-employed population who are unemployed from those who are not economically active.

There are many difficulties and subtleties in the application of these seemingly simple rules. Perhaps the most contentious is what is meant by "without work". Internationally this is accepted to be a person in the reference period, which is usually one week or one day before the survey question is posed, who did not work for pay or in-kind earning for **even one hour**.

Recognizing that too exclusive a focus on a single measure may distort the view of other developed nations in comparison with that of the United States, the US Bureau of Labour Statistics (BLS) has published since 1976 seven alternative measures for unemployment. Applying these definitions to nine countries Sorrentino (1993) found that the unemployment rate varied in 1989 for the US, for example, from 1.2% to 7.9%.

In developing countries, especially the poorer ones such as those in SSA, it is rare to find an adequate labour survey that follows precisely ILO labour standards. Data are normally available from ten-yearly-conducted censuses since resources are usually not available to carry out household surveys to map labour market statistics in the intervening periods. Consequently, almost without exception, unemployment figures that appear from time to time in the least developed countries are estimates of a very poor nature. To a certain extent this is understandable since the main labour issue in poor countries is normally under-employment, not the type of unemployment that results from the ILO's rather strict definition of unemployment. Yet, even rich countries do not always observe ILO definitions playing, for instance, with eliminating older people less than 65 who may not have a chance of a job from the register of the unemployed. EUROSTAT produces comparable unemployment estimates through taking the employment surveys from each member state and then re-calculating unemployment rates using the strict ILO definition.

Employment

According to the ILO's 1982 international definition of employment agreed at the Thirteenth International Conference of Labour Statisticians (Thirteenth ICLS) (see Hussmanns et. al., 1990), the "employed" comprise all persons above the age specified for measuring the economically active population, who, during a specified, brief period (one week or one day) were in the categories: (a) paid employment: "at work" or "with a job but not at work"; (b) self-employment: "at work": persons who performed some work for profit or family gain, in cash or in kind; (b) "with an enterprise but not at work". The international standards further specify that, for operational purposes, the notion of "some work' may be interpreted as work for *at least one hour*!

Under-employment

According to the 1966 ICLS resolution, under-employment exists "when a person's employment is inadequate, in relation to specified norms of alternative employment, account being taken of his occupational skill (training and work experience)". (Report of ICLS, 1966) Two principal forms of under-employment are distinguished: visible under-employment, reflecting an insufficiency in the volume of employment; and invisible under-employment, characterised by low income, underutilisation of skill, low productivity and other factors. The 1982 ICLS resolution weakened the emphasis on under-employment by noting that "for operational reasons the statistical measurement of under-employment may be limited to visible under-employment".

This means that visible under-employment is defined as a subcategory of employment and that there are three criteria for identifying, among persons in employment, those who are invisibly employed: (1) working less than normal duration; (2) doing so on an involuntary basis; and (3) seeking or being available for additional work during the reference period.

The 1982 ICLS noted that compared to visible under-employment, which is a statistical concept directly measurable by labour force and other surveys, invisible under-employment is "primarily an analytical concept reflecting misallocation of labour resources or a fundamental imbalance as between labour and other factors of production". To measure invisible under-employment, whether in respect of income, levels of skill or productivity, it is necessary to establish thresholds below which the income is considered abnormally low, the skill underutilised, or the productivity insufficient. However, the concept has not so far been endorsed by the ICLS because of the difficulty in defining international standards

Informal Sector Employment

The concept of the informal sector has played a growing role over the past three decades, in particular, in developing countries for its alleged role in absorbing vast numbers of unskilled labour in a dualistic economy. It is surprising, therefore, to note that no clear and generally accepted definition of the concept exists. The ILO, one of the first to introduce the concept in its Kenya Report in the early 1970s (ILO, 1972),has begun to take the concept even more seriously[24] in recent years as it attempts to develop a universally accepted definition and its measurement. Deliberations have taken place under the form of the Meeting of Experts on Labour Statistics (MELS). They note that one may broadly characterise it as the aggregate of activities that result from the need for generating one's own employment to earn a living because other sectors of the economy - agriculture, large modern firms and the public service - are unable to provide a sufficient number of adequate employment and income opportunities for a rapidly growing labour force and there

[24] The statistics department at least. King (1996) writes scathingly about current ILO research work vis à vis the informal sector, noting that the ILO now gives short shrift to the concept compared with its famous 1970s' Kenya Report.

are no - or only rudimentary - social benefits from the state to fall back on. It is by no means a marginal phenomenon - rough estimates put it at 300 million in developing countries (ILO, 1991). In developed countries, the labour surplus is smaller and social protection systems exist and therefore the informal sector that does exist (eg small-scale units outside the formal economy or services rendered by one household to another) is relatively small. Additionally, many activities exist in the black or concealed economy. In Central and Eastern Europe and some of the Southern European countries, the black economy is widespread.

The informal sector in developing countries typically consists of very small-scale units established and owned by self-employed persons either alone or in partnership with others. They often have very little capital, equipment, technical know-how or managerial skills, and use simple, labour-intensive technology. As a result, most of these units work at low levels of organisation, productivity and income. They tend to have little or no access to organised markets, credit institutions, formal education and training, or public services and amenities. The vast majority of these activities are legal in themselves, e.g. handicrafts, vehicle repair, taxi driving, food selling, etc., contrasted with criminal activities or illegal production, e.g theft, extortion, smuggling, production and distribution of drugs, prostitution etc.

On the other hand, governments often tolerate the existence of informal sector activities performed outside or at the fringe of laws and regulations because they lack the means to cover the whole economy with legislation or because they realise that many informal sector activities will one day become legal. Indeed, among informal sector entrepreneurs there is a desire to legalise their operations whenever possible, since this would enable them to have access to some institutional support, such as credit, or to the protection of the law in such matters as enforcement of contracts (Tokman, 1991).

In a draft resolution produced in its January meeting the 1993 ICLS adopted most of the recommendations for a definition made by MELS. (Report of ICLS, 1993) Much is still left to individual surveys or governments to decide on criteria, and further investigative work is currently underway at the ILO.

ANNEX 3D

EXERCISES IN LOCAL PLANNING

The purpose of this annex is to illustrate two exercises in local planning, efforts, in principle, dedicated to supplying useful information to those who require it at the local level: first, an early attempt by UNRISD and second, some of the procedures instituted by the ILO in their attempts to implement their "basic needs" approach to planning.

The UNRISD Programme

The programme to "measure real progress at the local level" instituted by the United Nations Research Institute for Social Development arose out of:

> dissatisfaction with the kind of information available to national planners and decision makers ... on progress in actually levels of living of the masses of the population ... The project aimed to improve the information systems for the benefit of development planners by making more use ... of information on conditions and changes at the local level ... and where locality studies are being carried out ... exploiting them more systematically for the use of planners. (quotes from Scott, 1973:1)

Leaving aside the undue emphasis on the planner's rather than the population's benefit, the interest of the programme here is the emphasis on:

> ...the formulation of new indicators (or the reformulation of old ones), ... the organisation that is required for the continuous collection, processing, analysis and presentation of information from the local level. (Scott, 1973:ii)

As they say, quoting a letter from Lenin to Gorki in July 1919, the rationale of using local level information is widely recognised. The problem is to implement it.

They proceed mainly by analysing cross-sectional data collected from interview or observation, cantering through a range of multivariate procedures which can be used to combine a wide range of items into a few policy-relevant indicators. But, because they recognise the importance of the choice of indicators and the difficulty of interpreting observed associations between them, they also draw attention to the possibility of using more complex methods to assess the objectives and aspirations of local populations.

More interesting, perhaps, are the range of methods they proposed for using data available at a local level, which would be easier and cheaper than national census, sample surveys and administrative files which have been subsequently taken up by the Rapid Appraisal advocates. These were:

- key persons reporting on selected and predetermined communal facilities and social arrangements;

- population and housing census enlarged in selected localities;

- making more use of administrative registers such as new housing starts, tax records at the local level;

- better use of sample surveys, especially of the conventional two-stage type, involving first selection of localities and then selection of respondents for interviewing.

Key to their approach was to organise interviewers to collect data on communal facilities; more intensive clustering to provide meaningful information for local aggregates.

On another level, local socio-economic observatories as have been set up in France: collating a wide variety of data about an area into a 'single' data base whilst the first ones were instituted for economic data, more recently, a similar network has been established for health data (Observatoires Regionaux de la Sante).

Techniques used by the ILO to Assess the Education-Employment Nexus at the Local Level

Those connected with the ILO have proposed a number of "techniques" (the inverted commas are used because they are really rather too derivative of other procedures to be separately named).

- calculating the "Basic Arithmetic" of Youth Employment simply by comparing the proportion of each age cohort with primary and secondary school qualifications with the proportions for whom there are modern sector 'jobs' available (Dore, 1976);

- relying on Key Informants (rather solemnly labelled as "a sociological - anthropological term") to provide household information on present economic activity, income levels, potential employment, future training and capital requirements in villages (McGranhan *et al*, 1982);

- local and regional analyses of the wage employment sector distinguished between subsistence self-employment and entrepreneurship, between casual, temporary labour, contracted and permanent workers in the formal sectors, apprentices, casuals and journeymen in the informal sector, and between cottage home industry, local firms and external enterprises (Kind, 1980).

It is clearly important:

> to move beyond catch-all phrases like self-employment and education for self-employment and become very specific" (Little, 1984, p.43).

But those categories she suggests (citing Gill, 1977), whilst possibly constituting a useful checklist, do not provide criteria for knowing **which learning needs** should be satisfied **in which socio-economic contexts**, let alone whether the education being provided might satisfy those learning needs.

These initial suggestions by the ILO – drawing on the UNRISD approaches - led to the subsequent development of the Rapid Rural Appraisal (RRA) family of techniques (again being solemn). These include using/seeking out existing information; identifying and learning from key informants; direct observations and asking questions about what is 'seen; guided interviews, group interviews with informal or selected groups. It means becoming a field researcher rather than a desk worker; the policy maker and planner learn from the villager; the villager becomes the subject not object of development. In particular, it highlights the importance of using existing information rather than commissioning a special survey (Myers 1985 cited in Thaker *et al*, 1988).

These procedures have proved useful in certain circumstances for local planners. Their usefulness will obviously depend on the extent to which the information collected answers questions which are seen as relevant by someone or some group. This raises the issue of who decided what information should be collected and what should be done with it when it has been collected. It should never be forgotten that control over local information can be an important means of consolidating and retaining power.

Radicalising Survey Methodology

Freire (1972:13) argued:

> There is no such thing as a *neutral* educational process. Education either functions as an instrument which is used to facilitate the integration of the younger generation into the logic of the present system and bring about conformity to it, *or* it becomes the 'practice of freedom', the means by which men and women deal critically and creatively with reality and discover how to participate in the transformation of their world.

On this basis, Freire developed his method of authentic-education for the oppressed whilst working with the peasants in north-eastern Brazil. His method consisted first of discovering the basic vocabulary and living conditions of the group to be taught; then, through group dialogue, the fundamental interests of the students. They then realise that they need to know more about the world before they can act consciously to control their own lives.
The problem with Freire's method (or similar prescriptions for radical education) is

that they only work *if people come to them*, whereas most people's experience of education (whether or not radical) is that the involvement and motivation by subjects of the learning process, which is crucial, cannot be assumed *ab initio*. They have to be 'drawn in' to education before they can be 'drawn out' by education. In fact, one main reason why such methods work (as indeed they do) is because of the ideals and enthusiasm of the committed, highly skilled and motivated educators who employ them. Indeed, it is arguable that someone like Freire would have been be a successful educator - in terms of raising consciousness so that people can participate in the 'transformation of their world' - with almost any method.

The problem, therefore, is to develop a low-key method of awakening people's interest in the first place (the subsequent level and nature of their motivation cannot, of course, be determined in advance); a tool which will focus people's attention on the issue, without imposing counter-productive discussion on those who do not see any problem or who have no hope of effecting any change. Any approach to uninterested and unmotivated people must, *inter alia*, examine their education experiences and the part these institutions have played in producing their immediate situation and particularly their understanding of, and reactions to, that situation. An early attempt was made by the author to use an interview/ questionnaire approach as this requires a very low level of involvement on the part of the respondent and yet introduces ideas to him/her.

The difficulty is, that when people are asked their reaction to the education they have received or that they would desire for their offspring, the purpose and content of the slab of education being offered are often indeterminate. Even if made precise, there is little incentive for people to reply because they sense that the final decisions will be made elsewhere.

We can go some way towards compiling data on needs in a non-alienating and non-exploitative fashion by involving a selection of the population at each stage. In a pilot study in Brighton, England, in 1974, the following procedure generated considerable discussion about the purpose and content of educational programmes:

- an informal discussion with groups of individuals from the projected population, eliciting the categories in which people perceive the reality of their own lives and possible futures;

- a more directed semi-structured interview with the same group of individuals about the relevance of present and possible educational careers to their own lives; and

- a self-completion questionnaire for the population designed so as to compare the purpose and results of present formal educational systems and other forms of socialisation, with the way in which they live their own life and their hopes for improvement in its quality. (taken from Carr-Hill,

1984b)

The experience of the questionnaire suggested that it was possible for respondents to be clear and coherent about what are desirable outcomes of all kinds from all forms of education in terms of attitudes, roles and skills - and, moreover, that *everyone is capable* of distinguishing between different educational contexts and their effects on these outcomes for themselves.

CONSOLIDATED REFERENCE LIST

R Abagi, O. (1997) "Education and Development of Human Capital", Draft Background Paper for Kenya National Human Development Report, UNDP

R Aiach P. and Carr-Hill R. (1989) "Inequalities in health: the country debate", in A.J. Fox (ed) Health Inequalities in European Countries, Aldershot, Gower for European Science Foundation, pp.19-49

R Aiach P, Carr-Hill R, Curtis S. and Illsley R (1988) Inégalités Sociales de Santé en France et en Grande Bretagne Paris, La Documentation Française pour INSERM

B Alkin M.C. (1988) "National Quality Indicators: a World View", Studies in Educational Evaluation Vol 14:1, pp.11-24

R Andrews F. and Withey S.B. (1976) *Social Indicators of Well-Being*, Plenum Press, New York

R Anthony, R.N. and Young, D.W. (1984) Management Control in Non-Profit Organisations (3rd Edition) Richard D. Irwin, Homeword

R Arnold M. (1969) "The twice revised code", in Smith P. and Summerfield G. (eds) Matthew Arnold and the Education of the New Order London, Cambridge University Press, pp. 157-198

R Ashenden D. (1988) An Odd Couple: Social Justice and Performance Indicators Working Paper, November, Social Justice, Standing- Council Victoria, Australia, referred to in Wyatt (1994)

R Baiman S. (1990) "Agency research in managerial accounting: a second look", Accounting, Organizations and Society Vol. 15:4, pp.341-371

R Bartelmus P., Stahmer C. and van Tongeren J. (1991) "Integrated environmental and economic accounting: framework for a SNA satellite system", Review of Income and Wealth, June, pp.111-148

R Baster, N. (1985) "Social Indicator Research: some issues and debates", in Hilhorst, J.G.M. and Klatter, M., Social Development in the Third World: level of living indicators and social planning. London: Croom Helm, pp.23-24

R Bauer R.A. (1966) Social Indicators, Cambridge, Mass, MIT Press

R Bennell P. (1996) "Rates of Return to Education: Does the Conventional Pattern Prevail in Sub-Saharan Africa?", World Development, Vol 24:1, pp183-99

R Bennell P. (1996) "Using and Abusing Rates of Return: A Critique of the World Bank's 1995 Education Sector Review", International Journal of Educational Development, Vol. 16:3, pp.235-248

R Bibeau, R. (1989). L'éducation non-formelle au Québec, Paris, International Institute for Educational Planning (IIEP Research Report No. 78)

R Bishop A. (1997) "Evaluation of Mathematics components of SIDA Programmes" Chapter VII, in Portuguese in R. Carr-Hill, R. Flavell *et al* (1997) Programmas de Apoio ao Sector de Educaçao em Moçambique, Education Division Documents, SIDA, Stockholm

R Blacker J.G.C. (1987) "Health Impacts of Family Planning" Health Policy and Planning, Vol. 2:3, pp.193-203, Oxford University Press

R Blaxter M. (1989) "A comparison of measures of inequality in morbidity", in Fox A.J. (ed) Inequalities in Health in European Countries, Gower, Farnborough, pp.199-230

R Blaxter M. (1985) "Self-definition of health status and consulting rates in primary care" Quarterly Journal of Social Affairs, Vol. 1:2, pp.131-171

R Blunt P. (1995) "Cultural relativism, 'good' governance and sustainable human development", Public Administration and Development, Vol. 15, pp.1-9

R Booth C. (1888) "Conditions and Occupations of the people of East London and Hackney" Journal of the Royal Statistical Society, Vol. 50:1

R Bottani N. and Delfau I. (1990) "Indicators of the Quality of education systems: an international perspective", International Journal of Education Research, Vol. 14:4, pp.321-408

R Bryk A. and Hermanson K.L. (1993) "Educational Indicator Systems: Observations on their structure, Interpretation and Use", in Review of Research in Education, Washington, D.C., AERA, pp.451-484

R Bulmer, (1990) "Problems of Theory and Measurement", Journal of Public Policy, Vol. 9:4, pp.407-12

R Caldwell, J.C. (1982) Theory of Fertility Decline, London, Academic Press

R Caldwell, J.C. (1993) "Health Transition: the cultural, social and behavioural determinants of health in the Third World", Social Science and Medicine, Vol. 36:2, pp.125-35

R Carley M. (1981) Social Measurement and Social Indicators: Issues of Policy and Theory London, George Allen and Unwin

R Carlsson, B.A. (1985). "The Potential of National Household Survey Programmes for Monitoring and Evaluating Primary Health Care in Developing Countries." World Health Statistics Quarterly, Vol. 38, pp.38-63

B Carr-Hill, R. (1984a). "The Political Choice of Social Indicators". Quality and Quantity, vol. 18, pp 173-91

R Carr-Hill, R. (1984b) "Radicalising Survey Methodology", Quality and Quantity, Vol. 18: pp. 275-92

R Carr-Hill R. (1989). "Background material for the workshop on QALYs: assumptions of the QALY procedure." Social Science and Medicine, Vol. 29, pp.469-77

R Carr-Hill, R. (1987) "When Is A Data Set Complete: a squirrel with a vacuum cleaner", Social Science and Medicine, Vol. 25:6, pp.753-764

B Carr-Hill R. (1988) The information base for the planning of the diversified educational field, Paris, International Institute for Educational Planning, Research Report No.68

B Carr-Hill R. (1990) Social Conditions in Sub-Saharan Africa, London, Macmillan

R Carr-Hill, R. (1991). "Allocating Resources to Health Care: is the QALY (Quality Adjusted Life Year) a Technological Solution to a Political Problem?", International Journal of Health Services, Vol. 21:2, pp. 351-63

R Carr-Hill R. (1992). "Health related quality of life measurement - Euro style". Health Policy, Vol. 20, pp.321-328

R Carr-Hill, R. (1995) Report on Mission for World Bank to Ivory Coast, 30/10/94-20/11/94, mimeo, Centre for Health Economics, University of York

R Carr-Hill R. and Magnussen O. (1973) Indicators of the Performance of Education Systems, Paris, OECD

R Carr-Hill R., Chengelele R, Kweka A and Rusimbi M. (1991) The functioning and effects of the Tanzanian Literacy Programme Paris, International Institute for Educational Planning, Research report No.68, Annex 5a

R Carrim, N. And Shalem, Yael, "School Effectiveness in South Africa", unpublished paper, Education Department, University of Witwatersrand, 1996

R Carron G. and Carr-Hill R. (1991) Non-Formal Education: Information and Planning Needs Paris, International Institute for Educational Planning

R Carron G. and Chau T.N. (1996) The Quality of Primary Schools in Different Development Contexts, Paris, International Institute for Educational Planning

R Carter N. (1989) "Performance indicators: 'backseat driving' or 'hands off' control?" Policy and Politics, Vol. 17, pp.131-138

R Cereseto S. & Waitzkin H. (1986) "Capitalism, Socialism and the Physical Quality of Life", International Journal of Health Services, Vol. 16, pp.643-658

R Cheng K-M. (1994) "Quality of Education as Perceived in Chinese Culture", p.74-84 in Takala *op cit*

R Clavijo S. (1992) "Variations on the basic needs yardstick", World Development, Vol.20, pp.1219-1223

R Committee of Privy Council on Education (1864) cited in Hogg (1990). *op cit.*

R Commonwealth Higher Education Management Service (1996). The Real World of Performance Indicators, prepared by Dorothy Davis, Director, International Students Division, IDP

R Cook T.J., Van Sant J., Stewart L and Adrian J. (1995) "Performance Measurement: Lessons Learned for Development Management", World Development, Vol. 23:8, pp.1303-1315

R Cooper, B., Lacey, D and Torrance, J. (1996) "Making Sense of Large Scale Evaluation Data: the Case of the Andhra Pradesh Primary Education Project" International Journal of Educational Development, Vol. 16:2, pp. 125-140

R Cornia, G.A., Jolly, R. and Stewart, F. (1987) Adjustment with a Human Face, Volume I, Protecting the Vulnerable and Promoting Growth, Oxford: Clarendon Press.

R Cornia, G.A., Jolly, R. and Stewart, F. (1988) Adjustment with a Human Face, Volume 2, Ten Case Studies, Oxford: Clarendon Press

R Crinnion J., Shotton J., and Carr-Hill R. (1999) "Literacy and Liquor Laws", mimeo, Institute of Education, submitted to World Development

R Crouch, L. (1996). "Notes on the national EMIS Proposal: From Data to Indicators", October, (internal paper), Department for National Education, South Africa.

B Culyer A.J. (1978) Measuring Health: Lessons from Ontario, University of Toronto Press, Toronto

R Culyer A.J. (1983) Health Indicators: an international study for the European Science Foundation, London: Martin Robertson

R Culyer, A.J. (1990) Health, health expenditures and equity, University of York, Centre for Health Economics, Discussion Paper, No. 83

R Cummings, W.K. and Riddell, A.R. (1992). Alternative Policies for the Finance Control and Delivery of Basic Education, Development Papers No 422ES, Harvard Institute for International Development.

R Dalin, P. (1994) How Schools Improve: An International Report, London: Cassell.

R Daly H. (1989) "Toward a Measure of Sustainable Social Net National Product", pp 8-9, in Y.J. Ahmad, S. El Serafy and E. Lutz (eds), Environmental Accounting for Sustainable Development, Washington, World Bank

R Daly H. and Cobb J.B. (1990) For the common good, Green Print

R Darling-Hammond, L. (1994) "Policy Uses and Indicators", Chapter 18 in OECD, 1994.

R Dasgupta, B. and Seers, D. (eds.) (1975). *Statistical Policy in Less Developed Countries*, Brighton, Institute of Development Studies, University of Sussex (IDS, No. 114).

R de Kadt, E. (1989) "Making Health Policy Management Intersectoral: issues of information analysis and use in less developed countries", Social Science and Medicine, Vol. 29: 4, pp. 503-14

R Department for International Development (1997) White Paper, London, HMSO Department for International Development

R Dore, R. (1976) The Diploma Disease: Education, Qualification and Development, London: George Allen and Unwin

R Dougherty, C.R.S. (1985) "Manpower Forecasting and Manpower-Development Planning in the United Kingdom", in R.V. Youdi and K. Hinchliffe (eds.), Forecasting Skilled Manpower Needs: The Experience of Eleven Countries, Comedi, Belgium, UNESCO

R Dowler, E.A., Payne, P.R., Young, O.S., Thomson, A.M. and Wheeler, E.F. (1982) "Nutritional Status Indicators: Interpretation and policy making role", Food Policy, May 1982, pp. 99-111

R Drummond, M. (1986) Resource Allocation Decisions in Health Care, Discussion Paper 21, Health Services Mangement Centre, University of Birmingham, 1986

R Economic Council of Japan (1972) Measuring Net National Welfare of Japan, Tokyo, Government Printing Office

R Edingbola, L.K., Walts, S.J., Kale, O.O., Smith, G.S. and Hopkins, D.R. (1986). "A Method of Rapid Assessment of the Distribution and Endemicity of Dracunculiasis in Nigeria". Social Science and Medicine, Vol. 23:6, pp. 555-8.

R Ekholm, M. and Karang, G. (1993) 'School-based evaluation as part of a national programme of evaluation: Early results', paper presented to the International Congress on School Effectiveness and Improvement (ICSE) 3-January, Noorkoping, Sweden

R Etzioni A. and Lehman E.W. (1967) "Some dangers in 'valid' social measurement", Annals of the American Academy of Political and Social Science, September, pp. 1-15

R EFILWC (European Foundation for the Improvement of Living and Working Conditions) (1992) Living Conditions in Urban Areas, Dublin, EFILWC

R Fanrian, A. (1986) "The perception of water quality among rural communities in South-Western Nigeria: Lessons for Planners", Water International, Vol. 11, pp. 169-74.

R Fergany, N. (1994) "Labour Market Returns to Education and Poverty in Egypt: Rewards or Punishments?", Research Notes, No. 07, Almishkat, Center for Research and Training, Zamalek, Cairo, March

R Fitz-Gibbon, C.T. (1990) (ed.) Performance Indicators Clevedon, Philadelphia, Multilingual Matters Ltd. (BERA Dialogues No.2)

R Flamholtz, E.G. (1983) "Accounting, budgeting and control systems in their organisational context: theoretical and empirical perspectives", Accounting, Organizations and Society, Vol. 8:2/3, pp.153-169

B Frackman, E. (1987) "Lessons to be learnt from a decade of discussions on performance indicators", International Journal of Institutional Management in Higher Education, Vol. 2:2, pp.149-62

B Franklin, J.J. (1988) "Performance Indicators in education: the participatory approach" in Wyatt and Ruby (eds.) *op cit.*

B Freire, P. (1972). Pedagogy of the Oppressed, Penguin: Harmondsworth

R Gallart, M.A. (1989) The diversified educational field in Argentina, Paris, International Institute for Educational Planning (IIEP Research Report No. 73)

R Gordon, G. (1979) "The Hungry Season in the Savanna of West Africa", presented to a conference on Rapid Rural Appraisal held at the Institute of Development Studies, University of Sussex, 4-7 December

R Government of Kenya (1996a) Eighth National Development Plan: 1997-2001 Nairobi

R Government of Kenya (1996b) Final Report of Workshop on Education and Training Objectives, Policies and Strategies, Utalii Hotel, Nairobi, Ministry of Education, November

R Government of Kenya (1997) Programme and Review and Forward Budget, 1997/98 - 1999/2000 Nairobi, February

R Gudex, C. and Kind, P. (1988) The QALY Toolkit, Discussion Paper 38, Centre for Health Economics, University of York

B Hanushek E. (1995) World Bank Research Observer, Vol 10

R Harloe M and Martens M (1984) "Comparative Housing Research", Journal of Social Policy, Vol. 13:3, pp. 255-277

R Hawes H. and Stephens D.J. (1992) Questions of Quality: Primary Education and Development, Essex, Longman

R Hicks and Streeton, P. (1979). "Indicators of development: the search for a basic needs yardstick", World Development, Vol. 7:6, pp.567-80

R Hirani Ratcliffe Development Consultants (1994) Economic Appraisal of Education projects?: Outputs, Outcomes and Indicators. A Review on Behalf of the ODA, April 1994

R Hofstede, G. (1981) "Management control of public and not-for-profit activities", Accounting, Organisations and Society, Vol. 6:3, pp.156-162

R Hogg, G.W. (1990) "Great Performance Indicators of the Past" Chapter 2 in Fitz-Gibbon, *op.cit.*

R Hogwood, B. W. and Gunn, L.A. (1984). Policy Analysis for the Real World. Oxford University Press: Oxford

R Hopkins, M. (1985). The MACBETH model for forecasting the impact of employment strategies

R Hopkins, M. (1999) The Planetary Bargain: Corporate Social Responsibility. London: Macmillan

R Hoppers, W. (1994) "Questioning the School - What Room for the Local Perspective" pp. 41-66, in Takala (ed.) *op. cit.*

R Hopwood, A.G. (1972). "An Empirical Study of the Role of Accounting Data in Performance Evaluation" Empirical Research in Accounting (Supplement to Journal of Accounting Research, pp.156-182

R Hueting, R. (1991) "Correcting National Income for Environmental Losses: A Practical Solution for a Theoretical Dilemma", in R. Costanza (ed.), Ecological Economics, Columbia University Press, pp. 194-213

R Humana, C. (1986) World Human Rights Guide.

B Hunter, D.J. (1996) Evidence-based Medicine: the Last of the Panaceas?, paper presented at London Medical Sociology Group Conference The Role of 'Consumers'/Users in Evidence Based Health Care, Wednesday 13th March 1996

R Hussmanns, R., Mehran, F. and Verma, V. (1990) "Survey of economically active population, employment, unemployment and underemployment", Geneva, ILO

R Illsley R. and Mitchell R. (1984) Low Birth Weight: a Medical, Psychological and Sociological Study. London: Wiley

R ILO (1991) The dilemma of the informal sector, Report of the Director General (Part I), International Labour Conference, 78th Session, Geneva

R ILO (1972) Employment, Incomes and equality: A Strategy for increasing productive employment in Kenya, ILO, Geneva

R ILO (1976) Employment Growth and Basic Needs: A One World Problem, ILO, Geneva

R Improving Educational Quality Project (1994) "Improving Educational Quality Project" (IEQ) South Africa, "Proceedings of a Seminar on Effective Schools, Effective Classrooms", Occasional Paper 1, May

R Innes, J.E. (1990) "Disappointments and Legacies of Social Indicators", Journal of Public Policy, Vol. 9:4, pp.429-432

R International Labour Office (1966) Report of the Fifth International Conference of Labour Statisticians, Geneva, ILO

R International Labour Office (1976) Employment, Growth and Basic Needs, Geneva, ILO

R International Labour Office (1982) Report of the Thirteenth International Conference of Labour Statisticians, Geneva, ILO

R International Labour Office (1984) Employment Planning , Santiago, PREALC

R International Labour Review (1994) Special Issue: Competitiveness, Equity and Skills, Vol. 133: 2

R Jansen, J.D (1995) Effective Schools, Vol. 31: 2, pp. 181

R Johannsen, S. (1976) Towards a Theory of Social Reporting. Stockholm: Swedish Institute for Social Research

R Johnston, D.F. (1988) "Towards a Comprehensive 'Quality of Life' Index", Social Indicators Research, Vol. 20, pp.473-496

R Jones, C.S. (1986) 'Universities, on becoming what they are not', Financial Accountability and Management, Vol. 2:2, pp.107-119

R Kalache, A. (1986) "Ageing in Developing Countries: are we meeting the challenge?" Health Policy and Planning, Vol. 1:2, pp. 171-175

R Kavanos, P. and Mossialos, E. (1996) The Methodology of International Comparisons of Health Care Expenditures, LSE Health Discussion Paper No.3.

R Kenis, I. (1979) "Effects of goal characteristics on managerial attitudes and performance", The Accounting Review, Vol. 54, pp.707-721

R Khaniya, T.R. (1997) Teacher support through resource centres: the Nepalese case, Paris, International Institute for Educational Planning (Trends in School Supervision Series).

R Kilele, A and Sinclair, C (1991) "Economic aspects of the informal sector in Kenya", Geneva, International Labour Office, mimeo., May

R Kind, P. and Rosser, R., (1978) "A scale of valuations of states of illness: is there a social consensus?" International Journal: of Epidemiology, Vol. 7, pp.347-358

R King, K. (1996) "Microenterprises and the Macroeconomic Environment: Revisiting Kenya's Informal Sector (Jua Kali) against the background of the formal globalising economy", International Journal of Educational Development, Vol.16; 4, pp.417-426

R Lauglo, J. (1996) "Banking on Education and the Uses of Research: A Critique of World Bank Priorities and Strategies for Education", International Journal of Educational Development, Vol. 16:3, pp.221-233

R Levin, H.M. and Lockheed, M.E. (1993) Effective Schools in Developing Countries. London: Falmer Press

R Lind, N.C. (1992) "Some thoughts on the HDI", Social Indicators Research, Vol.27, pp.89-101

R Lind, N.C. (1993) "A Compound Index of National Development", Social Indicators Research, Vol.28, pp. 267-284

R Little, A. (1984) " Education, earning and productivity - the eternal triangle", in J. Oxenham (ed.), Education Versus Qualifications. London: George Allen and Unwin

R Lourie, S. (1987) Are Consequences of Adjustment Policies on Education Measurable? International Institute for Educational Planning (UNESCO) Paris

R MacCormack, C.P. (1988) "Health and the Social Power of Women," Social Science and Medicine, Vol. 26: 7, pp 677-83

R Maslow, A. (1987) Motivation and Personality, (3rd Edition). New York: Harper and Row

R Mason, W. (1979) Guidelines for the Development of Employment and Manpower Information Programmes in Developing Countries: A Practical Guide. Geneva: ILO

R McDonnell, L.M. (1989) "The Policy Context", in Shavelson R.J., McDonnell L.M. and Oakes J. (eds.) Indicators for Monitoring Mathematics and Science Education: A Sourcebook. Rand Corporation, Santa Monica, CA, USA

R McEwan, M. and Hau Chow (1991) "Issues in Implementing Indicators Systems", Alberta Journal of Education Research, Vol. 37:1, pp. 65-86

R McGranahan, D.V., Pizarro, E. and Richard, C. (1985). Measurement and Analysis of Socio-Economic Development. Geneva: UNRISD

R McGranahan, D.V., Richard-Proust, C., Sovani, N.V. and Subramanian, M. (1972) Contents and Measurement of Socio-Economic Development. New York: Praeger for UNRISD

R McRae, D. (1990) "Policy Indicators: A Continuing and Needed Field", Journal of Public Policy, Vol. 9:4, pp. 437-8

R Meek, L. (1995) Insights from the Literature on the Use of Performance Indicators in Higher Education: an international comparison. Paper given at Stellenbosch, South Africa in September 1995 on National Information Systems and Performance Indicators

R Merchant, K. (1990) "The effect of management controls on data manipulation and management myopia", Accounting, Organizations and Society, Vol. 15:4, pp.297-313

R Miles, I. (1985) Social Indicators for Human Development. London: Pinter, for United Nations University

R Mills, A. "Decentralisation and accountability in the health sector from an international perspective: what are the choices?" Public Administration and Development, Vol. 14, 281-292

R Morris, M.D. (1979). Measuring the Conditions of the World's Poor : The Physical Quality of Life, New York

R Mosley, W.H. and Chen, L.C. (1984) Child Survival: strategies for research. Cambridge: Cambridge University Press

R Murray, C.J.L. (1987) "A Critical Review of International Mortality Data". Social Science and Medicine, Vol. 25:7, pp. 773-81

R Murray, C.J.L. (1991) Development Data Constraints and the HDI, UNRISD

R Namibia Ministry of Education and Culture, Florida State University, and Harvard University (1994) How Much Do Namibia's Children Learn in School?: Findings from the 1992 National Learner Baseline Assessment. Windhoek: Ministry of Education and Culture

R National Commission on Excellence in Education (1983) A Nation at Risk, Department of Health, Education and Welfare, Washington D.C.

B NEIDA, Network for Educational Innovations and Development in Africa, Education and Productive Work in Africa, UNESCO/BREDA/NEIDA, Dakar, n.d. (c.1979).

R National Foundation for Research into Human Resources Development (1982). Annex 5E, p3

R Nordhaus W.D. and Tobin J. (1973) "Is growth obsolete?" in M. Moss (ed.), The measurement of economic and social performance, NBER

R Nuttall D.L. (1992) "The Functions and Limitation of International Education indicators", in OECD, *op cit.*, pp.13-24

R Nuttall D.L. (1994) "Choosing Indicators", in Riley and Nuttall, *op. cit.*, pp.17-40

R Oakes, J. (1986) Education Indicators: A Guide for Policy Makers. Rand Corporation, Santa Monica, CA.

R Oketch, H O (1995) "Education and Training for the Informal Sector", in Fiona Leach (ed.) Education and Training for the Informal Sector, Vol.2. (Education Research Serial 11), Overseas Development Administration, March

R Oni, G.A. (1988) "Child mortality in a Nigerian city: its levels and socio-economic differentials", Social Science and Medicine, Vol 27:6, pp. 607-14

R OECD (Organisation for Economic Cooperation and Development), (1970) List of Social Concerns. Paris: *OECD*

R OECD (Organisation for Economic Cooperation and Development), (1973) Indicators of Performance of Education Systems. Paris: OECD

R OECD (Organisation for Economic Cooperation and Development), (1974) A Framework for Educational Indicators to Guide Government Decisions. OECD: Paris (document for general distribution).

R OECD Organisation of Economic Cooperation and Development (1976) Measuring Social Well-Being: A Progress Report of the Development of Social Indicators. Paris: OECD

R OECD Organisation of Economic Cooperation and Development (various years since 1993) Education at a Glance. Paris: OECD

R OECD, Organisation for Economic Cooperation and Development (1992) The OECD International Education Indicators: A Framework for Analysis. Paris, OECD

R OECD, Organisation for Economic Cooperation and Development (1994) Making Education Count: Developing and Using International Indicators, Centre for Educational Research and Innovation. Paris: OECD

R Otley, D.T. (1978) "Budget use and managerial performance." Journal of Accounting Research, Vol. 16, pp122-149

R ODA, Overseas Development Administration (1991a) Planning Development Projects, Internal Document

R ODA, Overseas Development Administration (1991b) Office Instructions : Project Reporting, Monitoring and Reviews Volume II: G1

R ODA, Overseas Development Administration (1993) Aid to Education: Strategy for 1993 and Beyond, London, Overseas Development Administration

R ODA, Overseas Development Administration (1995) "Terms of Reference for Evaluation and New Project Identification", SPRED, November

R ODA, Overseas Development Administration (1995) Internal Document, 'Defining Objectives and Measuring Performance in Education Activities', December

R ODA, Overseas Development Administration (1996) SPRED I Evaluation: The Government of Kenya, Ministry of Education, Evaluation Report, Strengthening of Primary Education (SPRED) Project, Hirani Ratcliffe Development Consultants

R ODA, Overseas Development Administration (1996) Measuring Performance in the Education Sector, Internal Document

R ODA, Overseas Development Administration(1996) Project Monitoring, Reports and Reviews, Office Instructions, Vol. II: G1, London, ODA, 27 May 1996

R ODA, Overseas Development Administration (1996) A Guide to Appraisal, Design, Monitoring, Management and Impact Assessment of Health and Population Projects, Health and Population Division

B Oulai, D. and Carr-Hill, R.A. (1994) The Impact of HIV/AIDS upon Education, Paris, International Institute of Educational Planning, Report of an IIEP Seminar

R Pateman, T. (1968) Counter Course, Harmondsworth, Penguin

R Petty, W. (1690) Political arithmetik, or a discourse concerning the extent and value of lands, people, buildings. Reprinted in C.H. Hull (1989) (ed.), Economic Works of Sir William Petty. Cambridge Union Press.

R Plomp, T. and Loxly, W. (1992) 'IEA and the Quality of Education in Developing Countries', in Vedder, P. (ed.) Measuring the Quality of Education. Amsterdam: Swets and Zeitlinger

R Report of International Conference of Labour Statisticians (ICLS) (1993), Geneva, ILO

R Report of International Conference of Labour Statisticians (ICLS) (1966), Geneva, ILO

R Republic of South Africa, Department of Education, National EMIS Steering Committee (1996a) A National Education management Information System (EMIS) for South Africa, Appendices to a Proposal presented at the National Workshop on 25-27 September

R Republic of South Africa, Department of Education, National EMIS Steering Committee (1996b) A National Education Management Information System (EMIS for South Africa, Proposal presented at the National Workshop on 25-27 September

R Republic of South Africa, Department of Education, National EMIS Steering Committee (1996c) Report on EMIS Study Tours to Twenty Countries, July

R Republic of South Africa, Department of Education (1996d) Changing Management to Manage Change in Education, Report of the Task Team on Education Management Development, December

R Republic of South Africa, Department of Education (1996e) Government Gazette (Vol. 368, No. 16987) General Notice, Notice 130 of 1996 "Education White Paper 2", Pretoria, February 14

R Republic of South Africa, Department of Education (1996f) Government Gazette (Vol. 370, No. 17118), "National Education Policy Act 1996" 24 April.

R Republic of South Africa, Department of Education (1996g) Government Gazette (Vol. 377, No. 17579), "South African Schools Act, 1996", 15 November

R Republic of South Africa, Department of Education (1996h) School Register of Needs Survey

R Republic of South Africa, Department of Education (1997a) Annual Survey for Schools: 22 April 1997

R Republic of South Africa, Department of Education (1997b) Curriculum 2005, April 1997

R Richards, P. (1994) "Issues in Manpower Analysis", in P. Richards and R. Amjad (eds.) New Approaches to Manpower Planning Analysis, Geneva, ILO

R Richter, L. (1989) "Upgrading Labour Market Information in Developing Countries: Problems, Progress and Prospects", Geneva, ILO

R Riddell, A.R. (1997) "Assessing Designs for School Effectiveness Research and School Improvement in Developing Countries", Comparative Education Review, Vol. 41:2, pp. 178-205

R Riley, K. A. and Nuttall, D.L. (1994) Measuring Quality: Education Indicators - United Kingdom and International Indicators. London: Falmer Press

R Riley, K. A. and Nuttall, D.L. (1994) 'Measuring Performance: National Contexts and Local Realities', in Riley and Nuttall, *op. cit.*, pp.122-131

R Ross, K. and Mahlck, L. (1990) Planning the Quality of Education: The Collection and Use of Data for Informed Decision-making. Oxford: Pergamon Press

R Ruby, A. (1989) 'Why invest in performance indicators development?' in Ruby, A . *et al* Reporting Educational Progress: Performance Indicators in Education. Bulletin 11, Monograph Series Australian Conference of Directors General of Education, Sydney

R Ruby, A. (1992) "Do Common Values Produce Common Indicators", in OECD, *op cit.*, pp.77-82

R Ruby, A. (1994) "Education Indicators: Officials, Ministers and the Demand for Information", in Riley and Nuttall, *op. cit.*

R Runciman, J. (1887) Schools and Scholars. London: Chatto

R Ruta, D.A., Garrett, A.M., Long, M., Russell, I.T. and MacDonald, L.M. (1994) "A New Approach to the measurement of the quality of life: the patient generated index", Medical Care, Vol. 32, pp.1109-1126

B Scheerens, J. (1992) "Process indicators of School Functioning", in OECD (1992), *op. cit.*, pp. 53-76

R Schumacher, E. (1973) Small is Beautiful: Economics as if People Mattered. London: Blond and Briggs

R Scott, W. with Argalis, H and McGranahan, D.V. (1973) The Measurement of Real Progress at the Local Level. Geneva: UNRISD

R Seers, D. (1975) "The political economy of national accounting", in A. Cairncross and M. Puri (eds.), Employment, Income Distribution and Development Strategy: Problems of the Developing Countries. London: Macmillan

R Seers, D. (1975) "A possible unified core system of social and demographic statistics for analysing, planning and monitoring social development" in Dasgupta and Seers (eds.), *op. cit.*

R Seers, D. (1979) "The meaning of development, with a postscript", in Lehmann, D. (ed.) (1979) Development Theory: Four Critical Studies. Bournemouth: Cass

R Seers, D. (1983) The Political Economy of Nationalism. Oxford: Oxford University Press

R Selden, R.W. (1994) "How Indicators have been Used in USA", in Riley and Nuttall, *op. cit.*, pp. 41-48

R Sheehan, G. and Hopkins, M. (1979) Basic Needs Performance: An Analysis of Some International Data. Geneva: International Labour Office

R Sheldon, E.B. and Moore, W.E. (1968) Indicators of Social Change. New York: Russell Sage Foundation

R Sigurdsson, S. and Schweitzer, E. (1994) The Use of Sectoral and Project Performance Indicators in Bank Financed Education Opportunities, World Bank, Education and Social Policy Department

R Sigurdsson, S. and Schweitzer, E. (1995) Performance Indicators in Bank-Financed Education Operations, Second Edition, Human Development Department, Washington, D.C., World Bank

R Sing, M.G. (1990) Performance Indicators in Education Australia, Deakin University, (ED 433/733 Management of Resources in School)

R Sizer, J., Bormans, R. and Spee, A. (1992) The role of performance indicators in higher education. Higher Education, Vol. 24: 2, pp. 133-155

R Smith, P. (1993) 'Outcome-Related Performance Indicators and Organisational Control in the Public Sector', British Journal of Management, Vol. 4, pp.135-51

R Sorrentino, C. (1993) "International comparisons of unemployment indicators", Monthly Labor Review, Vol. 116:3

R Stephens, D. (1991) 'The Quality of Primary Education in Developing Countries: who defines and who decides', Comparative Education, Vol. 27:2, pp. 223-233

R Sterky, G. and Millander, L. (eds.). (1978) Birth-Weight Distribution - An Indicator of Social Development, Stockholm, SAREC Report No. R:2

R Stern, J.D. and Hall, G (1987) 'Education Development at the Federal Level' CES, OERI, USA Department of Education, April

R Stewart, F. (1987) Planning for Basic Needs, London: Macmillan

R Stone, R. (1972) A System of Social and Demographic Statistics, New York, United Nations Statistical Office

R Takala T. (1994) Quality of Education in the Context of Culture in Developing Countries. Helsinki: Tampere University

R Tanner, J.M. (1982) "The potential of auxological data for monitoring economic and social well being", Social Science History, Vol. 6:4

R Thaker, S.B., Gibson-Parrish, R., Trowbridge, F.L. and Surveillance Coordinating Group (1988) "A Method for Evaluating Systems of Epidemiological Surveillance", World Health Statistics Quarterly, Vol. 41, pp.11-18

R Tokman, V.E. (1991) "The informal sector in Latin America: From underground to legality", in G. Standing and V.E. Tokman (eds.), Towards social adjustment: Labour Market issues in structural adjustment. Geneva: ILO

R UNESCO (1979) Final Report of the Regional Seminar on the Applicability of Socio-Economic Indicators for Analysis and Planning in Africa. Paris: UNESCO.

R United Nations Development Programme (1990) Human Development Report. Oxford: Oxford University Press

R UNRISD, 1964 Measurement of Real Progress at the Local Level. Geneva

R UNRISD (1977) Measurement of Real Progress at the Local Level. An Overview, Geneva, UNRISD/77/C.25

R United Nations Educational, Scientific and Cultural Organisation (1990) Education for All, Jomtien Conference on Education for All

B Uphoff, N. (1992) "Monitoring and Evaluation of Popular Participation in World Bank Assisted Projects", in Bhatnagar, B. and Williams, A. C. (eds.) Participatory Development and the World Bank, World Bank Discussion Paper 183

R van Herpen, M. (1992) "Conceptual Models in Use for Education Indicators", in OECD (1992)

R Walsh, K. (1994) "Quality, Surveillance and Performance Measurement", in Riley and Nuttall, *op. cit.*, pp.49-68

R Wapenhans Report (1990) Portfolio Management Task Force Report Washington, World Bank

R Weinstein, M.C. and Stason, W.B. (1977) "Foundations of cost-effectiveness analysis for health and medical practices", New England Journal of Medicine 296, 7165-7231

R Wilcox, B. (1990) "Performance Indicators: The Search for the Holy Grail of Quality", in Fitz-Gibbon (ed.), *op. cit.*, pp.31-40

R Wiles, P. (1971) "Crisis Prediction", Annals of the American Academy of Political and Social Sciences, January

R Williams, A. (1985) "The Economics of Coronary Artery By Pass Grafting", British Medical Journal, Vol. 291, pp.326-329.

R Wisner B. (1988) Power and Need in Africa: Basic Human Needs and Development Policies. London: Earthscan

R World Bank (1994) World Development Report 1994. Washington: World Bank

R World Bank (1996) Performance Monitoring Indicators: a handbook for task managers. Operational Policy Department, Washington, World Bank

R World Bank (1997a) Education Aide Memoire, Washington, D.C., March 5

R World Bank (1997b) Welfare Monitoring Survey I and II, Washington D.C., World Bank

R Wright, S. (1985) Age, Sex and Health: A Summary of Findings from the York Health Evaluation Survey. University of York, Centre for Health Economics, Discussion Paper No. 15

R Wunsch, G. (1983). "Maternal and Child Health in the Developing Countries: Problems of Data Collection", World Health Statistics Quarterly, Vol. 36, pp. 62-71

R Wyatt, T. and Ruby, A. (1988) 'Indicators in education'. Sydney: Australian Conference of Directors General of Education

R Wyatt, T. and Ruby, A. (eds.) (1990) Education Indicators for Quality, Accountability and Better Practice. Sydney: Australian Conference of Directors of Education

R Wyatt, T. (1994) "Education Indicators: A Review of the Literature", in OECD (1994), *op. cit.*, pp. 99-121

R Zolotas, X. (1981) Economic growth and declining social welfare. New York University Press.

R (1998) Evaluation in Primary Education – a Handbook for getting started. New Delhi: DFID India.

R (1999) Assessing Impact. London: DFID (Education Research Series).

DFID Department for International Development

Department For International Development
94 Victoria Street
London SW1E 5JL

ISBN: 1 86192 294 6